JOHANNES KEPLER
UNIVERSITÄT LINZ

Eingereicht von
DI Thalhammer Andreas

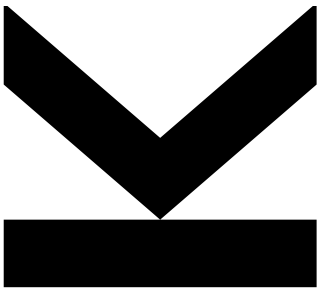
Angefertigt am
Doktoratskolleg Computational Mathematics, Institut für Stochastik

Erstbeurteilerin
Univ.-Prof. Dr. Evelyn Buckwar

Zweitbeurteiler
Prof. Gabriel Lord, PhD

Juni 2017

Numerical methods for stochastic partial differential equations: Analysis of stability and efficiency



Dissertation

zur Erlangung des akademischen Grades

Doktor der Technischen Wissenschaften

im Doktoratsstudium

Technische Wissenschaften

JOHANNES KEPLER
UNIVERSITÄT LINZ
Altenbergerstraße 69
4040 Linz, Österreich
www.jku.at
DVR 0093696

Abstract

This cumulative thesis contains several contributions to the numerical analysis of stochastic partial differential equations. The main focus lies on investigating and improving numerical methods with respect to the qualitative properties *stability* and *efficiency*. Here the term stability of a numerical method denotes a measure for the approximation quality of the considered numerical method based on fixed refinement parameters in space and time or on a finite number of independent realisations for Monte Carlo estimators. In contrast, the efficiency of a numerical method for approximating the solution process of SPDEs measures the computational work needed to obtain a certain accuracy. Besides the separate discussion of these two qualitative properties, the emphasis is laid on investigating their interplay and their practical relevance for numerical experiments.

This thesis consists of an introductory essay followed by 4 chapters based on 4 scientific articles. The purpose of the introduction is to provide an overview of relevant results on SPDEs and their approximations as well as discussing the existing literature on stability theory and efficiency investigations of numerical methods for stochastic differential equations.

In Chapter 2 we investigate the performance of Monte Carlo methods for linear stochastic differential equations with an asymptotically almost surely stable, but mean-square unstable equilibrium solution. It is illustrated that under this specific stability setting standard Monte Carlo estimators fail to reproduce the qualitative behaviour of the second (or higher) moment(s) of the solution process. As a remedy an importance sampling technique focusing on the simulation of rarely occurring realisations of the solution process is proposed and numerically tested.

In Chapter 3 we develop importance sampling techniques for SPDEs based on an infinite-dimensional version of the well-known Girsanov transformation. An optimality result that provides the existence of a measure transformation, for which the Monte Carlo error vanishes completely, is used as guidance for constructing measure transformations that can be easily implemented.

In Chapter 4 a structural mean-square stability theory for approximations of SPDEs is developed. For this we extend well-known mean-square stability results for approximations of finite-dimensional SDEs to an abstract tensor product-space setting and we derive necessary and sufficient conditions for the asymptotic mean-square stability of the zero solution of approximations of linear SPDEs. For a comparative study of numerical methods we investigate various combinations of rational approximations (of the semigroup) with ei-

ther Maruyama or Milstein time integration schemes. Furthermore, results connecting the stability properties of the zero solution of the SPDE and of its numerical approximations are derived.

In Chapter 5 we combine space-time multigrid techniques for deterministic partial differential equations with multilevel Monte Carlo methods for stochastic differential equations with additive noise. This coupling provides a new class of algorithms that are fully parallelisable, i.e., they can be computed in parallel with respect to space, time and probability.

Extensive numerical experiments in Chapters 2-5 illustrate the theoretical results.

Zusammenfassung

Diese kumulative Dissertation beinhaltet mehrere Beiträge zur numerischen Analysis von stochastischen partiellen Differentialgleichungen (im Folgenden als SPDEs abgekürzt). Der Hauptfokus richtet sich auf das Untersuchen und das Verbessern von numerischen Methoden für SPDEs bezüglich deren *Stabilität* und *Effizienz*. Wir bezeichnen hier unter Stabilität ein Maß für die Approximationsgüte eines numerischen Verfahrens unter Berücksichtigung endlicher Diskretisierungsparameter oder – im Falle von Monte Carlo Methoden – von Simulationen, die auf einer endlichen Anzahl an unabhängigen Realisierungen von den betrachteten Zufallsvariablen basieren. Unter Effizienz hingegen verstehen wir ein Maß für den zu leistenden Rechenaufwand, um eine bestimmte Genauigkeit zu erzielen. Neben der gesonderten Betrachtung dieser beiden qualitativen Eigenschaften liegt der Schwerpunkt der Arbeit auf dem Zusammenspiel und der praktischen Relevanz der untersuchten Eigenschaften für numerische Experimente.

Diese Dissertation besteht aus einem einleitenden Kapitel gefolgt von 4 weiteren Kapiteln, die auf 4 wissenschaftlichen Arbeiten basieren. Das Ziel der Einleitung ist es eine Übersicht über relevante Resultate sowohl für SPDEs und deren Approximationen als auch für die Stabilitätstheorie und für Effizienzuntersuchungen von numerischen Verfahren zu geben.

Wir untersuchen im Kapitel 2 die Performance von Monte Carlo Methoden für lineare, stochastische Differentialgleichungen, deren Nulllösung asymptotisch fast sicher stabil ist, aber instabil im Quadratmittel. Es wird gezeigt, dass in dieser speziellen Stabilitätskonfiguration Monte Carlo Schätzer das zweite (oder höhere) Moment(e) nur unzureichend approximieren können. Als Lösung entwickeln wir eine Importance Sampling Methode, mit der selten vorkommende Realisierungen des Lösungsprozesses simuliert werden können.

Im Kapitel 3 entwickeln wir Importance Sampling Methoden für SPDEs, die auf einer unendlich-dimensionalen Version des bekannten Girsanov Theorems beruhen. Ein Optimalitätsresultat, welches die Existenz einer Maßtransformation beweist, für die der Monte Carlo Fehler vollständig verschwindet, dient dabei als Konstruktionsanleitung für Maßtransformationen, die einfach implementiert werden können.

Im Kapitel 4 wird eine strukturelle Methode zur Untersuchung von Quadratmittel-Stabilität für Approximationen von SPDEs entwickelt. Um dies zu erreichen werden bekannte endlichdimensionale Resultate in einer abstrakten Tensor-Produktformulierung erweitert. Im Zuge einer vergleichenden Studie von numerischen Verfahren untersuchen wir Methoden, die auf verschiedenen Kombinationen von rationalen Approximationen mit Maruyama- oder

Milstein-Verfahren beruhen. Weiters werden Resultate, die die Stabilitätseigenschaften der Nulllösung der SPDE selbst und von deren Approximationen beschreiben, daraus abgeleitet.

Im Kapitel 5 kombinieren wir Mehrgitter-Verfahren in Raum und Zeit für deterministische partielle Differentialgleichungen mit Multilevel Monte Carlo-Verfahren für stochastische Differentialgleichungen mit additivem Rauschen. Diese Kopplung liefert einen neuen, bezüglich Raum, Zeit und Wahrscheinlichkeit vollständig parallelisierbaren Algorithmus.

Ausführliche numerische Experimente veranschaulichen in den Kapiteln 2-5 die theoretischen Resultate.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe. Die vorliegende Dissertation ist mit dem elektronisch übermittelten Textdokument identisch.

Linz, Juni 2017 Thalhammer Andreas

Acknowledgement

First, I would like to thank my PhD advisor Evelyn Buckwar for offering me the opportunity to work in the interesting research field of numerical analysis of SPDEs. Without her valuable help and her impressive library this thesis would have not been possible in such a short time. I would also like to thank Prof. Gabriel Lord for his willingness to be the external evaluator of my thesis.

Then, I want to thank my colleagues from the Institute of Stochastics, Markus Ableidinger, Harald Hinterleitner, Valentin Sturm, Massimiliano Tamborrino and Irene Tubikanec for offering me a nice and productive working environment. Many thanks also go to my colleagues from the Doctoral Program *Computational Mathematics* (DK), among which I would like to especially thank Simon Hubmer for his support and the fruitful discussions that we had during the last years.

Furthermore, I wish to express my gratitude to my co-authors, Annika Lang and Andreas Petersson, for their hospitality during my stay at Chalmers University of Technology, Gothenburg, and for the inspiring collaboration, which was for sure one of the most important ingredients for finishing this PhD thesis. As co-authors I would also like to thank here Markus Ableidinger and Martin Neumüller for the interesting and fruitful collaborations.

Finally, I would also like to express my deepest gratitude to my family for their valuable support. In particular I want to thank my girlfriend, Elisabeth, for her encouragement over the last years.

This thesis is based on research funded by the Austrian Science Fund (FWF): grant no. W1214-N15, project DK14.

x

Contents

1 Introduction

Over the last decades the importance of models describing time-dependent systems under random influences has gained a lot of attention across numerous fields of science. Typically such dynamical systems can be represented by stochastic differential equations (SDEs), where the driving noise process is modelled as a stochastic process with non-smooth paths with respect to time. Since explicit representations of solutions to SDEs can only be found in rather few cases, the development of efficient and stable numerical methods for SDEs has become an important and rapidly growing field of research.

In this thesis we mainly consider approximations of stochastic partial differential equations (SPDEs) attaining values in an infinite-dimensional Hilbert space H. Hence – besides the approximation of the solution process in time – one has to consider a discretisation with respect to the underlying infinite-dimensional state space H. In addition if one is interested in the expectation of a functional applied to the solution process, then a standard (but in general computationally expensive) method is to use Monte Carlo estimators based on a finite number of independent numerical trajectories of the solution process.

The majority of contributions to the numerical analysis of SPDEs is devoted to the strong and weak convergence properties of the considered numerical method, e.g. the investigation of the asymptotic behaviour of the numerical method as the spatial or temporal refinement parameters tend to 0, see e.g. [43, 49] for an overview. However, one has to be aware of the fact that for the actual implementation of numerical experiments one is essentially limited to non-zero spatial and temporal refinement parameters and to a finite number of independent realisations of the solution process for Monte Carlo simulations. For this reason, we investigate in this thesis qualitative properties of numerical methods using fixed refinement parameters for infinite-dimensional SDEs. The main focus of this thesis is on two specific qualitative aspects of numerical methods for SDEs, namely *stability* and *efficiency*.

For examining the stability properties of numerical methods the underlying question is how well does the numerical method approximate the qualitative properties of the analytical solution. This thesis covers various issues on stochastic stability ranging from investigating the asymptotic behaviour (with respect to time) of analytical solution trajectories and their approximations to the analysis of the effects of rare events on Monte Carlo estimators. For this reason, the term *stability of a numerical method* can be understood in the context of this thesis as a measure for the approximation quality based on fixed refinement parameters or on a finite number of independent realisations for Monte Carlo estimators.

The *efficiency* of a numerical method for approximating the solution of SDEs has to be

understood as a measure for the computational work needed to obtain a certain accuracy. Apparently in cases, where the expectation of a functional applied to the solution process is approximated by using Monte Carlo simulations, it is important to constantly improve the performance of the used algorithms in order to enhance the accuracy of numerical experiments within reasonable computational cost. For this, multilevel algorithms or purely probabilistic approaches such as e.g. importance sampling are proposed in the literature to limit the computational effort of numerical experiments. However, at this point the intrinsic interplay of stability and efficiency becomes evident, since such improved algorithms typically require (extended) stability properties of the considered numerical methods. For instance, if we consider the multilevel Monte Carlo estimator, it is clear that an appropriate numerical method approximating the level-wise Monte Carlo estimators has to be stable enough on each of the considered refinement levels - especially on the coarsest ones. Hence, besides the separated discussion of the two qualitative properties stability and efficiency, we are also investigating essential issues concerning the relationship and the interplay of stability, efficiency and convergence of numerical methods in this thesis.

The outline of this thesis is as follows: In the remaining part of Chapter 1 we set up the framework, which serves as the foundation for the results presented in Chapters 2-5. The main focus is on providing an overview of the scientific landscape of investigating stability and efficiency of numerical methods for S(P)DEs and on illustrating, where the main contributions of this thesis are. For this, we recall basic definitions and methods of stochastic stability theory and discuss various aspects of enhancing Monte Carlo simulations by multilevel algorithms, importance sampling techniques and by parallelisation.

In Chapter 2 we discuss the effects of different notions of stochastic stability on Monte Carlo simulations. More specifically, we examine the performance of Monte Carlo methods estimating the second moment of SDEs, for which the zero solution is asymptotically almost surely stable, but mean-square unstable. In this specific stability setting, standard Monte Carlo estimators do not reproduce the correct qualitative behaviour of the second moment of the solution process. In order to improve the performance of standard Monte Carlo estimators we propose an importance sampling technique that allows to sample rarely occurring realisations of the considered random variables, with which the qualitative behaviour of the second moment can be simulated more reliably.

In Chapter 3 we construct a class of importance sampling techniques for SDEs in infinite dimensions by using a version of the well-known Girsanov transformation for Hilbert space-valued stochastic processes. By this we gain that the resulting importance sampling technique is independent of the spatial discretisation technique. As a guidance to construct useful and implementable measure transformations we derive a (theoretically) optimal measure transformation, for which the Monte Carlo error vanishes completely. Besides the theoretical considerations the emphasis of this chapter is laid on discussing several implementation and approximation issues, under which convergence properties of the numerical method used for approximating the trajectories of the mild solution process are preserved.

In Chapter 4 we investigate mean-square stability properties of numerical methods for

approximating mild solutions of SPDEs. For this, we introduce an abstract tensor product space-valued framework, in which we find necessary and sufficient conditions for the asymptotic mean-square stability of the zero solution of the considered numerical method. Based on this, we perform a comparative mean-square stability analysis of various numerical methods ranging from Maruyama- and Milstein-type approximation schemes in combination with various rational approximations of the underlying semigroup. For backward Euler schemes we provide sufficient conditions for the simultaneous mean-square stability of the zero solution of the SPDE and of the corresponding approximation without any restriction of the involved refinement parameters.

In Chapter 5 we combine parallelisable space-time multigrid methods for deterministic time-dependent partial differential equations with multilevel Monte Carlo methods for SDEs with additive noise. Instead of applying the backward Euler scheme sequentially for every time step, the basic idea for the considered space-time method is to solve a large linear system at once, for which a parallelisable multigrid algorithm is constructed that inherits the space-time hierarchy of the multilevel Monte Carlo method. Overall this results in a fully parallelisable algorithm with respect to space, time and probability. As model problems we study in finite dimensions the Ornstein–Uhlenbeck process and in infinite dimensions the stochastic heat equation.

1 Stochastic differential equations in infinite dimensions

Let $[0,T], T > 0$, be the time horizon of interest and let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$ be a filtered probability space with a filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$ satisfying the "usual conditions" in the sense of [48, Definition 2.1.11] or [26, Section 2.1.2], i.e., complete and right continuous. In this thesis we consider semilinear stochastic differential equations (SDEs) attaining values in a separable Hilbert space $(H, \langle\cdot,\cdot\rangle_H, \|\cdot\|_H)$ given by

$$\mathrm{d}X(t) = AX(t) + F(X(t))\mathrm{d}t + G(X(t))\mathrm{d}W(t) \quad \text{for } t \in [0,T] \tag{1.1}$$

with initial value $X(0) = X_0 \in L^2(\Omega; H)$, where

$$L^2(\Omega; H) = \{v : \Omega \to H \mid v \text{ is strongly measureable, } \|v\|^2_{L^2(\Omega;H)} := \mathbb{E}[\|v\|^2_H] < \infty\}.$$

The linear operator $A : \mathcal{D}(A) \subset H \to H$ is assumed to be the generator of a strongly continuous semigroup $(S(t), t \in [0,T])$. Typically in this thesis, the operator A is an unbounded differential operator which gives rise to calling Equation (1.1) a *stochastic partial differential equation* (SPDE).

The driving noise process $W = (W(t), t \in [0,T])$ is a stochastic process attaining values in a separable Hilbert space $(U, \langle\cdot,\cdot\rangle_U, \|\cdot\|_U)$. Furthermore, we assume that $F : H \to H$ is a (possibly) non-linear, but sufficiently smooth operator and that for all $x \in H$ the diffusion operator $G(x)$ is a linear operator mapping elements from U to the solution space H. Concerning the diffusion operator we distinguish between the following two cases:

- If G is a non-constant mapping from H to $L(U;H)$ (the space of bounded linear operators mapping from U to H), then SPDE (1.1) is called *multiplicative*.
- In contrast if $G(x) \equiv B \in L(U;H)$ for all $x \in H$, then SPDE (1.1) is called *additive*.

Note that the above setting generalises the notion of finite-dimensional stochastic ordinary differential equations (SODEs) as the following example shows:

Example 1.1. *Choose $H = \mathbb{R}^d$ and $U = \mathbb{R}^m$. Then we consider the SODE*

$$\begin{aligned} \mathrm{d}X(t) &= f(X(t))\,\mathrm{d}t + g(X(t))\,\mathrm{d}\beta(t), \qquad t \in [0,T], \\ X(0) &= x_0 \in \mathbb{R}^d, \end{aligned} \tag{1.2}$$

where $\beta(t) = (\beta_1(t), \dots, \beta_m(t))^T$ is an m-dimensional Brownian motion, $f : \mathbb{R}^d \to \mathbb{R}^d$, and for all $x \in \mathbb{R}^d$, $g(x)$ is a linear mapping from $\mathbb{R}^m$ to $\mathbb{R}^d$. Furthermore, the operator A corresponds to the zero matrix in this example.

In order to guarantee the existence of (mild) solutions to SPDE (1.1), we have to specify additional properties of the involved coefficients in Equation (1.1). For this, we start by discussing the driving noise process W in Section 1.1. Afterwards, we show in Section 1.2 that under specific assumptions on the regularity of the coefficients F, G, X_0 and A, there exists a unique mild solution to SPDE (1.1). Finally, we discuss in Section 1.3 numerical approximations of the mild solution process.

1.1 Hilbert space-valued stochastic processes

Since we discuss various types of stochastic processes in this thesis, we provide in this section a short overview of stochastic processes attaining values in a separable Hilbert space $(U, \langle\cdot,\cdot\rangle_U, \|\cdot\|_U)$. We start by recapitulating standard results on Q-Wiener processes that are treated in Chapters 2-5. Afterwards we note how these results can be extended to processes that are *rough* in space by discussing properties of the cylindrical Wiener process. Finally, we briefly discuss generalisations of U-valued Wiener processes such as square-integrable martingales and Lévy processes (that are both treated in Chapter 4).

Q-Wiener processes

As a first motivation we introduce the Q-Wiener process in finite dimensions. For this consider an m-dimensional Brownian motion $\beta(t)$ as in Example 1.1, for which we denote by $\Delta\beta^{t,s} = \beta(t) - \beta(s), 0 \le s < t$, its independent increments. For an arbitrary matrix $\Sigma \in \mathbb{R}^{d\times m}$ we define the finite-dimensional Q-Wiener process by $W(t) = \Sigma\beta(t), t \in [0,T]$. Note that the covariance of the increments $\Delta W^{t,s} = \Sigma\Delta\beta^{t,s}$, $0 \le s < t$, of the finite-

dimensional Q-Wiener process satisfies

$$\begin{aligned}\mathrm{Cov}[\Delta W^{t,s}, \Delta W^{t,s}] &= \mathbb{E}[\Delta W^{t,s}(\Delta W^{t,s})^T] \\ &= \mathbb{E}[(\Sigma\,\Delta\beta^{t,s})(\Sigma\,\Delta\beta^{t,s})^T] \\ &= \mathbb{E}[\Sigma\,\Delta\beta^{t,s}\,(\Delta\beta^{t,s})^T\,\Sigma^T] \\ &= \Sigma\mathbb{E}[\,\Delta\beta^{t,s}\,(\Delta\beta^{t,s})^T]\Sigma^T = (t-s)\,(\Sigma\Sigma^T),\end{aligned}$$

i.e., the covariance $(t-s)Q := (t-s)\Sigma\Sigma^T$ is a symmetric and non-negative definite matrix.

Analogously to the finite-dimensional process, one defines the infinite-dimensional Q-Wiener process in the following way:

Definition 1.2. Let $Q \in L(U)$ be a self-adjoint, non-negative definite operator of trace class, i.e., $\mathrm{tr}(Q) < \infty$. Then, $W = (W(t), t \in [0,T])$ is called a Q-Wiener process if and only if

1. $W(0) = 0 \quad \mathbb{P}$-a.s.,
2. for any $0 \le r \le s \le t$ it holds that the increments $W(t) - W(s)$ are independent of $W(r)$,
3. W has $\mathbb{P}$-a.s. continuous paths,
4. the increments satisfy $\mathbb{P} \circ (W(t) - W(s))^{-1} \sim \mathcal{N}(0, (t-s)Q)$, where $\mathcal{N}(0,(t-s)Q)$ denotes a U-valued Gaussian measure.

For further details on Hilbert space-valued Gaussian measures we refer to [22, 59, 26, 43].

Remark 1.3. In this introduction we use the term *Wiener process* for Hilbert space-valued stochastic processes, whereas *Brownian motion* refers explicitly to a real-valued, finite-dimensional stochastic process. For being able to subsequently distinguish between these two notions of stochastic processes, we use $(W(t), t \in [0,T])$ for U-valued Wiener processes and $(\beta(t), t \in [0,T)])$ for a finite-dimensional, real-valued Brownian motion.

Under the above assumptions it is shown e.g. in [59, Proposition 2.1.10] or [49, Theorem 10.7] that the Q-Wiener process attains the representation

$$W(t) = \sum_{k=1}^{\infty} \sqrt{\mu_k} f_k \beta_k(t), \tag{1.3}$$

where $(f_k, k \in \mathbb{N})$ is an orthonormal basis of U consisting of eigenfunctions of the covariance operator Q and $(\beta_k(t), k \in \mathbb{N})$ is a sequence of independent, real-valued Brownian motions. Note that the series (1.3) converges in $L^2(\Omega; U)$ due to the trace class property of Q. Subsequently we call the representation of W from Equation (1.3) *the* Karhunen–Loève expansion of the Q-Wiener process W.

For the construction of the stochastic integral with respect to a Q-Wiener process, the space of Hilbert–Schmidt operators plays an important role. For any linear operator B

mapping from U to H, the Hilbert–Schmidt norm is given by

$$\|B\|^2_{L_{HS}(U,H)} = \sum_{k=1}^{\infty} \|B\varphi_k\|^2_H,$$

which is independent of the choice of an orthonormal basis $(\varphi_k, k \in \mathbb{N})$ of U. Then, we obtain for the Q-Wiener increments $\Delta W^{t,s}, 0 \leq s < t$, and for any operator $B \in L(U;H)$

$$\begin{aligned}\|B\Delta W^{t,s}\|^2_{L^2(\Omega;H)} &= \mathbb{E}\left[\left\langle \sum_{k=1}^{\infty} \sqrt{\mu_k} B f_k \Delta\beta_k^{t,s}, \sum_{\ell=1}^{\infty} \sqrt{\mu_\ell} B f_\ell \Delta\beta_\ell^{t,s} \right\rangle\right] \\ &= \sum_{k,\ell=1}^{\infty} \sqrt{\mu_k\mu_\ell}\, \mathbb{E}[\Delta\beta_k^{t,s}\Delta\beta_\ell^{t,s}]\langle Bf_k, Bf_\ell\rangle_U \\ &= (t-s)\sum_{k=1}^{\infty} \|B\sqrt{\mu_k} f_k\|^2_H = (t-s)\|BQ^{1/2}\|^2_{L_{HS}(U;H)},\end{aligned}$$

where the existence of $Q^{1/2}$ is guaranteed by the non-negative definiteness of Q. By introducing the notation $U_0 = Q^{1/2}(U)$ for the reproducing kernel Hilbert space $Q^{1/2}(U)$, we can conclude

$$\|B\Delta W^{t,s}\|^2_{L^2(\Omega;H)} = (t-s)\|B\|^2_{L_{HS}(U_0;H)}.$$

This norm is very important for the construction of the stochastic integral

$$\int_0^T G(X(s))\mathrm{d}W(s)$$

appearing in the SPDE (1.1), since a key ingredient for defining such a U-valued stochastic integral is the so-called Itô isometry: For this, let $(\phi(t), t \in [0,T])$ be a predictable and $\mathcal{F}_t$-adapted stochastic process. Then, we get that (see e.g. [26, Theorem 2.3])

$$\mathbb{E}\left[\left\| \int_0^T \phi(t)\mathrm{d}W(t)\right\|^2_H\right] = \mathbb{E}\left[\int_0^T \|\phi(t)\|^2_{L_{HS}(U_0;H)}\mathrm{d}t\right], \tag{1.4}$$

under the assumption that the right hand side is well-defined. For further details on the construction of the stochastic integral with respect to a Q-Wiener process we refer to the classical monographs [22, 26, 59].

Cylindrical Wiener processes

The results on Q-Wiener processes can be extended to stochastic processes with covariance operators $\tilde{Q}$ that are not of trace class. These infinite-dimensional Wiener processes are typically called *cylindrical* Wiener processes.

One can show the $L^2(\Omega;U)$-convergence of the expansion (1.3) of the Q-Wiener process by using the fact that $\mathrm{tr}\, Q < \infty$ guarantees the existence of a Hilbert-Schmidt embedding from $(U_0, \langle\cdot,\cdot\rangle_0)$ to $(U, \langle\cdot,\cdot\rangle_U)$. This cannot be shown in the presence of a non-trace

class covariance operator $\tilde{Q}$. However, as it is discussed in [59, Section 2.5.1], we can circumvent this problem by considering another (larger) Hilbert space $(U_1, \langle\cdot,\cdot\rangle_{U_1})$ and a Hilbert-Schmidt embedding $J : U_0 \to U_1$. Then, by [59, Proposition 2.5.2], we get that $Q_1 = JJ^* \in L(U_1)$ is a non-negative definite and symmetric operator with finite trace and that the series

$$W(t) = \sum_{k=1}^{\infty} \beta_k(t) J f_k$$

defines a Q_1-Wiener process on U_1. Note that by [59, Remark 2.5.1] the space U_1 and the embedding J can always be constructed by choosing $U_1 = U$ and for $u \in U_0$ define

$$J(u) = \sum_{k=1}^{\infty} a_k \langle u, f_k\rangle_0 f_k,$$

where $(a_k)_{k\in\mathbb{N}}$ is chosen such that $\sum_{k=1}^{\infty} a_k^2 < \infty$. Results on the construction of a stochastic integral with respect to a cylindrical Wiener process can be found e.g. in [26, Section 2.2.4].

Square-integrable martingales and Lévy processes

In Chapter 4 we investigate qualitative properties of numerical methods for stochastic partial differential equations driven by square-integrable martingales $(M(t), t \geq 0)$ that are càdlàg and adapted to the filtration $(\mathcal{F}_t, t \geq 0)$. For the construction of the corresponding stochastic integral we assume that there exists a symmetric, non-negative definite trace class operator $Q \in L(U)$ such that for all $0 \leq s \leq t$ it holds that

$$\langle\langle M, M\rangle\rangle_t - \langle\langle M, M\rangle\rangle_s \leq (t-s)Q,$$

where the double angle bracket process $(\langle\langle M, M\rangle\rangle_t, t \geq 0)$ is defined as

$$\langle\langle M, M\rangle\rangle_t = \int_0^t Q_s \,\mathrm{d}\, \langle M, M\rangle_s .$$

Here the single angle bracket process $(\langle M, M\rangle_t , t \geq 0)$ is the unique, increasing, predictable process from the Doob–Meyer decomposition and $(Q_t, t \geq 0)$ denotes the martingale covariance. Details on the construction of stochastic integrals with respect to U-valued square-integrable càdlàg martingales can be found in [58, Section 8.2] and [50, Section 3.6]. Note that these results generalise the case of Q-Wiener processes as the following example shows:

Example 1.4. *A U-valued Q-Wiener process is a $\mathbb{P}$-a.s. continuous, square-integrable martingale with $\langle W\rangle_t = t(\operatorname{tr} Q)$ and $\langle\langle W\rangle\rangle_t = tQ$.*

In this thesis we also consider in Chapter 4 square-integrable martingales called Lévy processes that are not necessarily $\mathbb{P}$-a.s. continuous:

Definition 1.5. A U-valued stochastic process $(L(t), t \geq 0)$ is called a Lévy process if

1 $L(0) = 0 \quad \mathbb{P}$-a.s.

2 L has independent and stationary increments,

3 L is a stochastically continuous process, i.e. for all $t \geq 0$ and for all $\varepsilon > 0$ it holds that

$$\lim_{s \to t} \mathbb{P}(\|L(s) - L(t)\|_U > \varepsilon) = 0.$$

Note that the martingale covariance $Q \equiv Q_s$ is stationary for Lévy processes and that L admits a Karhunen–Loève expansion

$$L(t) = \sum_{i=1}^{\infty} \sqrt{\mu_i} L_i(t) f_i,$$

where $(\mu_i, i \in \mathbb{N})$ are the eigenvalues of the covariance operator Q with respect to eigenfunctions $(f_i, i \in \mathbb{N})$ and $(L_i(t), t \geq 0)$ is a sequence of square-integrable, uncorrelated, real-valued Lévy processes.

Example 1.6. *Since $\mathbb{P}$-a.s. continuity implies stochastic continuity and the increments of a Q-Wiener process are by definition independent and stationary it holds that a Q-Wiener process is also a square-integrable Lévy process.*

1.2 Existence of solutions

The investigation of existence and uniqueness of strong solutions to stochastic ordinary differential equations has gained a lot of attention in the literature, see e.g. [4, 56, 60]. In contrast to the finite dimensional setting, the concept of strong solutions is too restrictive in the context of SPDEs, since this would in general require that the solution process satisfies $X(t) \in \mathcal{D}(-A)$ for all $t \in [0, T]$. However, the operator $-A$ typically fulfils the following assumption (see also [43, Assumption 2.13]):

Assumption 1.7. *The linear operator $-A : \mathcal{D}(-A) \subset H \to H$ is a densely defined, self-adjoint and positive definite operator with compact inverse.*

Under this assumption we get by results from [43, Appendix B] that the semigroup $(S(t), t \in [0, T])$ generated by A is analytic and that there exists an increasing, unbounded sequence $(\lambda_k, k \in \mathbb{N})$ of eigenvalues with corresponding eigenfunctions $(e_k, k \in \mathbb{N})$, i.e., $-Ae_k = \lambda_k e_k$. Thus, the restriction $X(t) \in \mathcal{D}(-A)$ for all $t \in [0, T]$ is in general not fulfilled since

$$\|AX(t)\|_H^2 = \sum_{k=1}^{\infty} \lambda_k^2 \langle X(t), e_k \rangle_H^2 < \infty$$

holds only under some restrictive regularity assumptions on X. We illustrate this by considering the one-dimensional stochastic heat equation in the following example:

Example 1.8. *Let $D = (0,1)$, $H = L^2(D)$ and $A = \Delta_x = \partial^2/\partial x^2$ with homogeneous Dirichlet boundary conditions. It is well-known that the eigenvalues of the Laplace operator $-A$ are given by $\lambda_j = \pi^2 j^2$ for $j \in \mathbb{N}$. Hence $\|Ax\|_H^2 < \infty$ if and only if $\langle x, e_j\rangle_H \simeq j^{-5/2-\varepsilon}$ for any $\varepsilon > 0$.*

Instead of strong solutions we investigate the existence and uniqueness of mild solutions of SPDE (1.1). For this, the stochastic process W is assumed to be a Q-Wiener process. A square-integrable mild solution is a predictable stochastic process $X = (X(t), t \in [0,T])$ satisfying

$$\sup_{t\in[0,T]} \mathbb{E}[\|X(t)\|_H^2] < \infty$$

and for all $t \in [0,T]$ the mild solution is given by

$$X(t) = S(t)X_0 + \int_0^t S(t-s)F(X(s))\mathrm{d}s + \int_0^t S(t-s)G(X(s))\mathrm{d}W(s) \qquad \mathbb{P}\text{-a.s.} \quad (1.5)$$

For results on the existence of such mild solutions we have to ensure sufficient regularity of the coefficients of SPDE (1.1). For this reason, we introduce the following assumption on the operators F and G:

Assumption 1.9. *(Linear growth conditions and Lipschitz continuity of F and G)*
Assume that F and G from SPDE (1.1) *satisfy*

1 that there exist constants $c_F, C_F > 0$ such that

$$\|F(\phi)\|_H \le c_F(1 + \|\phi\|_H),$$
$$\|F(\phi_1) - F(\phi_2)\|_H \le C_F\|\phi_1 - \phi_2\|_H$$

for all $\phi, \phi_1, \phi_2 \in H$ and

2 that there exist constants $c_G, C_G > 0$ such that

$$\|G(\phi)\|_{L_{HS}(U_0;H)} \le c_G(1 + \|\phi\|_H),$$
$$\|G(\phi_1) - G(\phi_2)\|_{L_{HS}(U_0;H)} \le C_G\|\phi_1 - \phi_2\|_H$$

for all $\phi, \phi_1, \phi_2 \in H$.

Furthermore, we have to impose additional regularity for the initial value:

Assumption 1.10. *The initial value satisfies $X_0 \in L^2(\Omega; \mathcal{D}((-A)^{1/2}))$.*

Under Assumptions 1.7, 1.9 and 1.10 the existence and (up to modifications) uniqueness of a square-integrable mild solution to SPDE (1.1) is guaranteed by Theorem 2.25 in [43]. Existence and (up to modifications) uniqueness of mild solutions of SPDEs driven by square-integrable Lévy processes can be found in [58, Section 9]. Further extensions of the above results to SPDEs with locally Lipschitz continuous operators F and G that satisfy the linear growth conditions from Assumption 1.9 can be found in [67].

Itô's formula for SPDEs

For SODEs (as considered in Example 1.1), the Itô formula (or also called Itô's lemma) is one of the central results in the finite-dimensional stochastic analysis: For $\varphi \in C^2(\mathbb{R}^d;\mathbb{R})$, i.e., a twice continuously differentiable function mapping from $\mathbb{R}^d$ to $\mathbb{R}$, the Itô formula is given by

$$\begin{aligned}\mathrm{d}\varphi\left(X(t)\right) = \Big[\nabla_x\varphi(X(t))\cdot f(X(t)) + \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2\varphi}{\partial x_i\partial x_j}(X(t))(g(X(t))g(X(t))^T)\Big]\mathrm{d}t \\ + \sum_{r=1}^{m}\nabla_x(X(t))\cdot g_r(X(t))\mathrm{d}\beta_r(t) \qquad \mathbb{P}\text{-a.s.}\end{aligned}$$

However, extending this result to infinite dimensions faces some limitations. In the case that $(X(t), t\in[0,T])$ denotes the strong solution of SPDE (1.1) and $\varphi \in \mathcal{C}_b^2(H;\mathbb{R})$, i.e., the mapping $\varphi : H \to \mathbb{R}$ has bounded first and second Fréchet derivatives (denoted by $D\varphi$ and $D^2\varphi$), the infinite-dimensional Itô formula can be written as, see e.g. in [22, 14],

$$\begin{aligned}\varphi(X(t)) = \Big[D\varphi(X(t))\left(AX(t)+F(X(t))\right) + \frac{1}{2}\sum_{j=1}^{\infty}D^2\varphi(X(t))(G(X(t))f_j, G(X(t))f_j)\Big]\mathrm{d}t \\ + D\varphi(X(t))G(X(t))\mathrm{d}W(t) \qquad \mathbb{P}\text{-a.s.}\end{aligned}$$

However, in many applications there does not exist a strong solution to SPDE (1.1) and thus, the expression on the right hand side is in general not well-defined. Hence, the straightforward extension of the finite-dimensional Itô formula to infinite-dimensional Hilbert space-valued equations is in general not possible.

There are a few important exceptions, where the infinite-dimensional Itô formula can be applied to SPDEs as the following example shows:

Example 1.11. *Consider the stochastic heat equation from Example 1.8 with a diffusion operator that satisfies for all $h \in H$ that $G(h) \in L_{HS}(U;H)$. Furthermore, we specify $\varphi(h) = \|h\|_H^2$. Then it is shown e.g. in [21] and the references therein that*

$$\begin{aligned}\|X(t)\|_H^2 = \Big[2\big\langle X(t), AX(t)+F(X(t))\big\rangle_H + \|G(X(t))\|_{L_{HS}(U;H)}\Big]\mathrm{d}t \\ + \big\langle X(t), G(X(t))\mathrm{d}W(t)\big\rangle_H \qquad \mathbb{P}\text{-a.s.}\end{aligned}$$

Note that results that are comparable to the finite-dimensional Itô formula are of high interest in the numerical analysis of SPDEs, since these results would offer to easily adapt a large number of proofs from the finite-dimensional stochastic analysis to the analysis of infinite-dimensional SPDEs.

There are already some promising attempts towards a useful infinite-dimensional version of Itô's formula available in the literature. Recently, a generalisation of the finite-dimensional Itô formula to mild solutions (1.5) has been considered in [21], where it has

been shown that

$$\begin{aligned}\varphi(X(t)) = \varphi\big(S(t)X_0\big) &\\ + \int_0^t D\varphi\big(S(t-s)X(s)\big)S(t-s)F(X(s)) &\\ + \frac{1}{2}\sum_{j=1}^{\infty} D^2\varphi\big(S(t-s)X(s)\big)\big(S(t-s)G(X(s))f_j, S(t-s)G(X(s))f_j\big)\mathrm{d}s &\\ + \int_0^t D\varphi\big(S(t-s)X(s)\big)S(t-s)G(X(s))\mathrm{d}W(s) \qquad \mathbb{P}\text{-a.s.}&\end{aligned}$$

1.3 Approximation of mild solutions

In this section we recapitulate some basic results on approximation techniques for Hilbert space-valued stochastic partial differential equations. An approximation of the mild solution (1.5) of SPDE (1.1) is a family of random variables denoted by $(X_h^k, k = 0, \dots, K)$ attaining values in a finite-dimensional subspace $V_h \subset H$ with $\dim(V_h) = N_h$. Here, $K \in \mathbb{N}$ denotes the number of elements of an equidistant partition Θ_K of the time interval $[0, T]$ defined by

$$\Theta_K = \{0 = t_0 < t_1 < \cdots < t_K = T\},$$

where $t_k = k\Delta t$ for a given time step size $\Delta t = T/K$. By using this notation, we denote by $X_h^k \in V_h$ an approximation of the mild solution $X(t_k)$ at a given time point $t_k \in \Theta_K$.

Note that for the numerical approximation of mild solutions of SPDEs, one has to consider at least three different types of approximation errors due to the approximation in space (with refinement parameter h), in time (with time step size Δt), and the approximation error in probability. The latter error results either from approximating the driving noise process by e.g. truncating the corresponding Karhunen–Loève expansion (1.3) or – in cases, where one is interested in the statistical properties of the mild solution process – from approximating the expectation of $\mathbb{E}[\varphi(X(T))]$ by using a Monte Carlo method.

There exists a rapidly growing literature on the convergence of approximations of semilinear SPDEs, where the asymptotic behaviour of an approximation scheme with respect to $h, \Delta t \to 0$ is discussed. In this thesis we focus on two different characterisations of stochastic convergence, i.e. *mean-square* and *weak* convergence. A numerical method approximating the mild solution (1.5) is called mean-square convergent if

$$\max_{t_k \in \Theta_K} \|X(t_k) - X_h^k\|_{L^2(\Omega;H)} \to 0$$

for $h, \Delta t \to 0$. Furthermore, a numerical method converges weakly to the mild solution (1.5) if for all $\varphi \in \mathcal{C}_b^2(H;\mathbb{R})$ it holds that

$$\max_{t_k \in \Theta_K} |\mathbb{E}[\varphi(X(t_k)) - \varphi(X_h^k)]| \to 0$$

for $h, \Delta t \to 0$. A very active field of research is the investigation of the mean-square and

weak convergence *orders* (or *rates*) of a numerical method. A numerical method is called mean-square convergent of order $\gamma > 0$ if there exists a constant $C > 0$ such that

$$\max_{t_k \in \Theta_K} \|X(t_k) - X_h^k\|_{L^2(\Omega;H)} \leq Ch^\gamma,$$

where we assume that there exists an appropriate coupling of all refinement parameters with respect to the spatial refinement parameter h, e.g., for Galerkin finite element methods combined with a time integration based on the backward Euler scheme, we obtain that there exist constants $C > 0$ and $\gamma > 0$ such that

$$\max_{t_k \in \Theta_K} \|X(t_k) - X_h^k\|_{L^2(\Omega;H)} \leq C(h^\gamma + \Delta t^{\gamma/2})$$

and hence, an appropriate coupling is given by $\Delta t \simeq h^2$.

Furthermore, a numerical method is called *weakly convergent of order* $\eta > 0$ if there exists for all $\varphi \in \mathcal{C}_b^2(H;\mathbb{R})$ a constant $C > 0$ such that

$$\max_{t_k \in \Theta_K} |\mathbb{E}[\varphi(X(t_k)) - \varphi(X_h^k)]| \leq Ch^\eta,$$

where we again assumed that all refinement are appropriately coupled with respect to h.

In the remaining parts of this section we discuss various methods and aspects for approximating the mild solution process (1.5). We start with the discussion of spatial discretisation techniques such as finite differences or Galerkin methods. Then we consider numerous time integration methods based on rational approximations, for which convergence results are provided. As third topic we discuss the error by truncating the Karhunen–Loève expansion (1.3) and fast simulation techniques for the increments of a Q-Wiener process. Finally, we conclude this section by investigating Monte Carlo estimators, for which we provide results on the approximation error.

Approximation in space

In this part we review basic results on spatial discretisation (or also called *semidiscretisation*) techniques for approximating the mild solution (1.5) in space. Here we discuss two different types of spatial approximation schemes, namely, *finite differences* and *Galerkin methods*.

For the discussion of the finite difference method we restrict ourselves for simplicity to the setting introduced in Example 1.8 (one-dimensional stochastic heat equation with homogeneous Dirichlet boundary conditions on $D = (0,1)$). Following [49], we first introduce an equidistant grid $\mathcal{T}_h = \{0 = x_0 < x_1 < \cdots < x_{N_h} = 1\}$, where $x_k = kh$ with spatial grid width $h = 1/N_h$. By Taylor's theorem we get for the interior grid points

$$\frac{\partial^2}{\partial x^2} X(t,x_j) = \frac{X(t,x_{j+1}) - 2X(t,x_j) + X(t,x_{j-1})}{h^2} + r_j(t), \quad j = 1,\ldots,N_h - 1,$$

for $t \in [0,T]$, where each of the remainder terms satisfy $r_j(t) = \mathcal{O}(h^2)$. By neglecting

these remainder terms we obtain the centered finite difference approximation of SPDE (1.1), which leads by defining

$$\mathbf{x}(t) = [X(t, x_1), \dots, X(t, x_{N_h-1})]^T,$$

to the following finite-dimensional SODE system

$$\begin{aligned} \mathrm{d}\mathbf{x}(t) &= [\mathbf{A}_h\mathbf{x}(t) + \mathbf{F}(\mathbf{x}(t))]\ \mathrm{d}t + \mathbf{G}(\mathbf{x}(t))\,\mathrm{d}\mathbf{W}(t) \quad \text{for } t \in [0, T], \\ \mathbf{x}(0) &= [X_0(x_1), \dots, X_0(x_{N_h-1})]^T, \end{aligned} \tag{1.6}$$

where

$$\begin{aligned} \mathbf{F}(\mathbf{x}(t)) &= [F(X(t, x_1)), \dots, F(X(t, x_{N_h-1}))]^T, \\ \mathbf{G}(\mathbf{x}(t)) &= [G(X(t, x_1)), \dots, G(X(t, x_{N_h-1}))]^T, \\ \mathbf{W}(t) &= [W(t, x_1), \dots, W(t, x_{N_h-1})]^T. \end{aligned}$$

The finite difference method has already been analysed and applied to semilinear SPDEs as given in (1.1), for which e.g. in [63] strong convergence of the finite difference method applied to SPDE (1.1) with additive noise is considered. Besides results on strong convergence there are also other topics treated in the literature such as e.g. the investigation of mean-square stability properties of a finite difference approximation in combination with a specific Milstein scheme in [61] or combining finite difference methods with multilevel Monte Carlo estimators in [30]. Furthermore, a detailled discussion on the finite difference method for the approximation of a cylindrical Wiener process can be found in [49].

In this thesis, the finite difference method is applied within a method-of-lines approach, i.e., we interpret the finite difference approximation in space as a finite-dimensional SODE system given in Equation (1.6). The major contribution of this thesis in the context of finite difference approximations is twofold: First, we perform in Chapter 2 a detailed mean-square and almost sure stability analysis of the finite difference approximation (1.6) of a linear version of the stochastic heat equation from Example 1.8, where we choose $F = 0$ and G being the Nemytskii operator induced by the linear function $\gamma(x) = \sigma x, \sigma \in \mathbb{R}$, see e.g. [43] for details. In particular, the impact of the parameter σ on the stability properties of the zero solution is examined. Second, we develop in Chapters 2 and 3 various importance sampling techniques for estimating $\mathbb{E}[\varphi(X(T))]$ based on finite difference approximations of the mild solution in order to reduce the Monte Carlo error.

For *Galerkin methods* we choose a sequence $(V_h, h \in (0, 1])$ of finite-dimensional subspaces of $\dot{H}^1 = \mathcal{D}((-A)^{1/2})$ indexed by a refinement parameter h, where for each $h \in (0, 1]$ the corresponding subspace V_h has dimension N_h. Furthermore, we denote by P_h the orthogonal projection onto V_h. Then, the semidiscrete problem is given by: Find

$X_h : [0,T] \to V_h$ such that

$$\mathrm{d}X_h(t) = [A_h X_h(t) + P_h F(X_h(t))]\,\mathrm{d}t + P_h G(X_h(t))\,\mathrm{d}W(t), \tag{1.7}$$
$$X_h(0) = P_h X_0. \tag{1.8}$$

Here, the linear operator $-A_h : V_h \to V_h$ is defined for each $v_h \in V_h$ by letting $-A_h v_h$ be the unique element of V_h such that

$$\langle -A_h v_h, w_h \rangle_H = \left\langle (-A)^{1/2} v_h, (-A)^{1/2} w_h \right\rangle_H$$

for all $w_h \in V_h$. This implies that $-A_h$ is a self-adjoint and positive definite operator on V_h and therefore $-A_h$ has a set of orthonormal eigenfunctions $(e_{h,i}, i = 1, \ldots, N_h)$ with corresponding positive, non-decreasing eigenvalues $(\lambda_{h,i}, i = 1, \ldots, N_h)$. In particular, this implies that $-A_h$ is the generator of a strongly continuous semigroup $(S_h(t), t \in [0,T])$. By the same arguments as for the infinite-dimensional problem, there exists a unique mild solution X_h for the semidiscrete problem (1.7), which $\mathbb{P}$-a.s. satisfies for $t \in [0,T]$

$$X_h(t) = S_h(t)P_h X_0 + \int_0^t S_h(t-s)P_h F(X_h(s))\,\mathrm{d}s + \int_0^t S_h(t-s)P_h G(X_h(s))\,\mathrm{d}W(s).$$

In this thesis Galerkin methods are used for the analysis and implementation of numerical experiments in Chapters 3-5, where the following two types of Galerkin approximation schemes are used:

Example 1.12. *Consider the operator $Au = \nabla \cdot (a(x)\nabla u) - c(x)u$ for all $x \in D$, where $D \subset \mathbb{R}^d, d = 1,2,3$, is a bounded, convex domain. For $d = 2$ the domain D is assumed to be polygonal and for $d = 3$ the domain D is polyhedral. Furthermore, if we consider Dirichlet boundary conditions, then let $a, c : D \to \mathbb{R}$ be sufficiently smooth and satisfy $a(x) \geq a_0 > 0$ $c(x) \geq 0$. If periodic boundary conditions are considered, then we have to additionally ensure that $c(x) \geq c_0 > 0$ in order to fulfil Assumption 1.7.*

(a) (Standard finite element method) Let $\mathcal{T}_h$ be a regular family of triangulations (see e.g. [13, 49]), where the refinement parameter h denotes the maximal mesh size of $\mathcal{T}_h$. Then V_h is defined as the space of globally continuous functions that are piecewise linear on the triangulation $\mathcal{T}_h$.

(b) (Spectral Galerkin method) Let $V_h = \mathrm{span}(e_1, \ldots, e_{N_h})$, where $(e_k, k = 1, \ldots, N_h)$ are the first N_h orthonormal eigenfunctions of the operator A.

Note that in numerical experiments the spectral Galerkin method is very effective if the operators A and Q share the same eigenfunctions.

Approximation in time

In this part we recall basic results for fully discrete approximations of the mild solution (1.5), i.e., an approximation in space by Galerkin methods and in time by using one-step time integration schemes with a fixed time step size $\Delta t > 0$.

Remark 1.13. Since the method-of-lines approach for the finite difference method leads to a system of SODEs and the resulting system of SODEs is mainly considered in the finite-dimensional context within this thesis (see Chapter 2) we omit here further discussions on a combination of finite differences and time integration methods. For more details on standard time integration methods for finite-dimensional SODE systems we refer the reader to [39, 55, 49] and we refer to [63, 61, 30] for details on time integration methods in combination with a finite difference approximation in space.

In this thesis we consider time integration techniques based on rational approximations of the underlying semigroup $(S(t), t \geq 0)$. For this we recall that a rational function $R : \mathbb{C} \to \mathbb{C}$ is called a rational approximation of the exponential function of order q if there exist constants $C, \delta > 0$ such that for all $z \in \mathbb{C}$ with $|z| < \delta$ it holds that

$$|R(z) - e^z| \leq C|z|^{q+1} \qquad \text{as } z \to 0. \tag{1.9}$$

By the definition of a rational function, there exist two polynomials r_d (denominator) and r_n (nominator) such that for all $z \in \mathbb{C}$

$$R(z) = \frac{r_{\mathsf{n}}(z)}{r_{\mathsf{d}}(z)}.$$

For the approximation of the semigroup $(S_h(t), t \in [0, T])$ generated by the operator $-A_h$, we consider the linear operator $R(\Delta t A_h) : V_h \to V_h$, which is defined for all $v_h \in V_h$ by

$$R(\Delta t A_h)v_h = \sum_{k=1}^{N_h} R(-\Delta t \lambda_{h,k})\langle v_h, e_{h,k}\rangle_H e_{h,k} = \sum_{k=1}^{N_h} \frac{r_{\mathsf{n}}(-\Delta t \lambda_{h,k})}{r_{\mathsf{d}}(-\Delta t \lambda_{h,k})}\langle v_h, e_{h,k}\rangle_H e_{h,k}.$$

An important class of rational approximations are the so-called *Padé approximants*, see e.g. [68]: For $\mu, \nu \in \mathbb{N}$ let

$$r_{\mathsf{n}}(z) = \sum_{j=0}^{\nu} \frac{(\mu+\nu-j)!\nu!}{(\mu+\nu)!j!(\nu-j)!} z^j \quad \text{and} \quad r_{\mathsf{d}}(z) = \sum_{j=0}^{\mu} \frac{(\mu+\nu-j)!\mu!}{(\mu+\nu)!j!(\mu-j)!}(-z)^j.$$

Note that the order q (see Equation (1.9)) of a Padé approximant satisfies $q = \mu + \nu$. The following example provides an overview of the most important rational approximations:

Example 1.14. *In this thesis we focus on the following three Padé approximants:*

- *The backward Euler scheme (Padé approximant with $\mu = 1$ and $\nu = 0$):*

$$R(z) = \frac{1}{1-z},$$

- *the Crank-Nicolson scheme (Padé approximant with $\mu = 1$ and $\nu = 1$):*

$$R(z) = \frac{1 + z/2}{1 - z/2},$$

- *and the forward Euler scheme (Padé approximant with $\mu = 0$ and $\nu = 1$):*

$$R(z) = 1 + z.$$

In this thesis we consider two different classes of approximation schemes for the stochastic integral (in combination with rational approximations of the semigroup): *Maruyama-type* and *Milstein-type* methods. Here, we start by considering the Galerkin Euler–Maruyama method, which is given by the recursion

$$\begin{aligned} X_h^{j+1} &= R(\Delta t A_h) X_h^j + r_d^{-1}(\Delta t A_h) \left(\Delta t P_h F(X_h^j) + P_h G(X_h^j) \Delta W^{j+1} \right), \\ X_h^0 &= P_h X_0, \end{aligned} \tag{1.10}$$

for $j \in \mathbb{N}_0$, where $\Delta W^{j+1} = W(t_{j+1}) - W(t_j)$. A commonly used member of this class of numerical approximation methods of the mild solution (1.5) is called the *backward* (also called *linearly implicit* or *semi-implicit*) *Euler–Maruyama* method given by the recursion

$$\begin{aligned} X_h^{j+1} &= X_h^j + \Delta t \left(A_h X_h^{j+1} + P_h F(X_h^j) \right) + P_h G(X_h^j) \Delta W^{j+1}, \qquad j \in \mathbb{N}_0 \\ X_h^0 &= P_h X_0. \end{aligned} \tag{1.11}$$

By using the representation of $R(z)$ for the backward Euler scheme from Example 1.14, the recursion (1.11) can be equivalently rewritten as

$$\begin{aligned} X_h^{j+1} &= R(\Delta t A_h) \left(X_h^j + \Delta t P_h F(X_h^j) + P_h G(X_h^j) \Delta W^{j+1} \right), \qquad j \in \mathbb{N}_0 \\ X_h^0 &= P_h X_0. \end{aligned} \tag{1.12}$$

In the literature, there are several results on the strong convergence of numerical approximations based on the backward Euler-Maruyama scheme in combination with Galerkin methods for the space discretisation, see e.g. [43, 6, 49]. If we consider the setting of Example 1.12 (stochastic heat equation with a semidiscretisation based on standard finite elements with respect to regular triangulations or spectral Galerkin methods) we obtain by [43, Theorem 3.14] that there exists a constant $C > 0$ independent of Δt and h such that

$$\max_{t_j \in \Theta_K} \|X_h^j - X(t_j)\|_{L^2(\Omega;H)} \leq C(h + \Delta t^{1/2}). \tag{1.13}$$

In order to obtain an appropriate coupling of the spatial refinement parameter h and the time step size Δt, we choose $\Delta t \simeq h^2$, which implies that we balance the error from the semidiscretisation and from the approximation in time such that Equation (1.13) reduces to

$$\max_{t_j \in \Theta_K} \|X_h^j - X(t_j)\|_{L^2(\Omega;H)} = \mathcal{O}(h).$$

It is well-known in the literature on the numerical analysis of finite-dimensional SODE systems that the weak convergence rate of the (backward) Euler–Maruyama method is in general twice the strong convergence rate (with respect to the time step size), see e.g. [39,

66]. Such results have only been proven in the literature on the numerical analysis of SPDEs for a few types of equations until now, since the lack of strong solutions does not allow to use similar analytical tools as in finite dimensions (e.g. Itô formula). In [2], Malliavin calculus has been used to prove a corresponding result for SPDEs with additive noise: Under the assumption that there exists a $\rho \in (0,1]$ such that $\|A^{(\rho-1)/2}\|_{L_{HS}(U_0;H)} < \infty$, then for all $\gamma \in [0,\rho)$ we get that

$$\max_{t_j \in \Theta_K} \|X_h^j - X(t_j)\|_{L^2(\Omega;H)} \leq C(h^\gamma + \Delta t^{\gamma/2}),$$
$$\max_{t_j \in \Theta_K} |\,\mathbb{E}[\varphi(X_h^j) - \varphi(X(t_j))]| \leq C(h^{2\gamma} + \Delta t^\gamma).$$

Note that in the general multiplicative noise case, the mean-square convergence order $1/2$ (with respect to the time step size) is already the best possible strong convergence rate that one can obtain by using numerical approximation schemes based on Maruyama-type approximation of the stochastic integral, see e.g. [19, 11]. In order to overcome this order barrier, extensions of the well-known finite-dimensional Milstein scheme to SPDEs are proposed in the literature, see e.g. [35, 5, 42, 27]. In Chapter 4 we qualitatively analyse the Milstein scheme from [5] for general Galerkin methods, which is given for $j \in \mathbb{N}_0$ by

$$\begin{aligned} X_h^{j+1} = R(\Delta t A_h) X_h^j + r_d^{-1}(\Delta t A_h)\left(\Delta t P_h F(X_h^j) + P_h G(X_h^j)\Delta W^{j+1}\right) \\ + \int_{t_j}^{t_{j+1}} r_d^{-1}(\Delta t A_h) P_h G\Big(\int_{t_j}^{s} G(X_h^j)\mathrm{d}W(r)\Big)\,\mathrm{d}W(s), \end{aligned} \tag{1.14}$$

where $X_h^0 = P_h X_0$. Then, by using the Karhunen–Loève expansion of the Q-Wiener process W we can rewrite the iterated stochastic integral in Equation (1.14) in terms of iterated stochastic integrals of the real-valued Brownian motions $(\beta_k, k \in \mathbb{N})$ from (1.3), i.e.,

$$\begin{aligned} &\int_{t_j}^{t_{j+1}} r_d^{-1}(\Delta t A_h) P_h G\Big(\int_{t_j}^{s} G(X_h^j)\,\mathrm{d}W(r)\Big)\,\mathrm{d}W(s) \\ &\quad = \sum_{k,\ell=1}^{\infty} \sqrt{\mu_k \mu_\ell}\, r_d^{-1}(\Delta t A_h) P_h G(G(X_h^j) f_k) f_\ell \int_{t_j}^{t_{j+1}} \int_{t_j}^{s} \mathrm{d}\beta_k(r)\,\mathrm{d}\beta_\ell(s). \end{aligned}$$

The scheme (1.14) turns out to have a higher mean-square convergence order than the Galerkin backward Euler–Maruyama method, since e.g. in the framework of the standard finite element method from Example 1.12 and additional regularity assumptions on the mild solution $(X(t), t \in [0,T])$, the Milstein scheme (1.14) satisfies that there exists a constant $C > 0$ such that (see [5])

$$\max_{t_k \in \Theta_K} \|X_h^j - X(t_j)\|_{L^2(\Omega;H)} \leq C(h^2 + \Delta t).$$

If G is a linear mapping from H to $L_{HS}(U_0;H)$ then the iterated stochastic integrals can be efficiently simulated if the following commutativity assumption is fulfilled:

Assumption 1.15. *([5, 35]) We assume that for all H-valued, $\mathcal{F}_t$-adapted stochastic pro-*

cesses $\chi = (\chi(t), t \geq 0)$ *and for all* $i, j \in \mathbb{N}$ *the diffusion operator satisfies*

$$G(G(\chi)f_j)f_i = G(G(\chi)f_i)f_j.$$

Under Assumption 1.15 the iterated stochastic integral can be represented by only using the increments of the real-valued Brownian motions, since it holds that

$$\begin{aligned}\int_{t_j}^{t_{j+1}} r_d^{-1}(\Delta t A_h) P_h G \left(\int_{t_j}^{s} G(X_h^j) \mathrm{d}W(r) \right) \mathrm{d}W(s) \\ = \frac{1}{2} \sum_{k,\ell=1}^{\infty} \sqrt{\mu_k \mu_\ell} r_d^{-1}(\Delta t A_h) P_h G(G(X_h^j) f_k) f_\ell (\Delta \beta_k^{j+1} \Delta \beta_\ell^{j+1} - \delta_{k,\ell} t),\end{aligned}$$

where $\Delta \beta_k^{j+1} = \beta_k(t_{j+1}) - \beta_k(t_j)$. However, Assumption 1.15 is not generally fulfilled as we can see by considering the following example:

Example 1.16. *We choose* $H = \mathbb{R}^d$ *and* $U = \mathbb{R}^2$*. We consider the linear SODE system*

$$\begin{aligned}\mathrm{d}X(t) &= G_1 X(t) \mathrm{d}\beta_1(t) + G_2 X(t) \mathrm{d}\beta_2(t), \\ X(0) &= X_0,\end{aligned}$$

where $G_1, G_2 \in \mathbb{R}^{d \times d}$ *and* $\beta_1(t), \beta_2(t)$ *are two independent real-valued Brownian motions. Then Assumption 1.15 is fulfilled if and only if* $G_1 G_2 = G_2 G_1$*. However, this assumption is clearly not fulfilled in general since the matrix multiplication is not commutative.*

If the diffusion operator does not satisfy the commutativity assumption, the computation of the iterated stochastic integrals becomes severely more difficult, since the simulation of the corresponding stochastic Lévy areas is required. In the literature there are efficient ways proposed to simulate these iterated stochastic integrals, see e.g. [23, 40, 69], however, in the case of semidiscretised SPDEs, the number of iterated stochastic integrals is typically coupled to the dimension of V_h (see below for discussions on approximating the stochastic process W) and for this reason a refinement in space leads to a larger number of iterated integrals that have to be simulated. Hence, the gain of a higher order of convergence by using the Milstein scheme faces a loss in the efficiency due to an increase in the computational complexity for operators G that do not satisfy Assumption 1.15.

Noise approximation

For SPDEs, for which the driving noise process attains values in an infinite-dimensional Hilbert space U, we also have to consider an appropriate noise approximation. As it has been pointed out in Section 1.1, the Q-Wiener process can be represented in terms of the series (1.3), which is not directly suitable for implementations, since the simulation of infinitely many standard Brownian motions would be required for numerical experiments. A standard way to solve this problem is to truncate the Karhunen–Loève expansion (1.3), see

[7, 49]: By choosing an appropriate truncation index $\kappa \in \mathbb{N}$ we define

$$W^{\kappa}(t) = \sum_{k=1}^{\kappa} \sqrt{\mu_k} f_k \beta_k(t).$$

For notational convenience we define $W^{c\kappa}(t) = W(t) - W^{\kappa}(t)$. Then we get for the truncation error in the mean-square sense that

$$\|W^{c\kappa(t)}\|^2_{L^2(\Omega;U)} = \Big\| \sum_{k=\kappa+1}^{\infty} \sqrt{\mu_k} f_k \beta_k(t) \Big\|^2_{L^2(\Omega;U)} = t \sum_{k=\kappa+1}^{\infty} \mu_k,$$

where we used Parseval's identity and $\mathbb{E}[\beta_k^2(t)] = t$ to obtain the last identity. For the stochastic integral we obtain the following truncation error

$$\begin{aligned} \Big\| \int_0^T \psi(s) \mathrm{d}W(s) - \int_0^T \psi(s) \mathrm{d}W^{\kappa}(s) \Big\|^2_{L^2(\Omega;H)} &= \Big\| \int_0^T \psi(s) \mathrm{d}W^{c\kappa}(s) \Big\|^2_{L^2(\Omega;H)} \\ &= \mathbb{E}\left[\int_0^T \|\psi(s)\|^2_{L_{HS}((Q^{c\kappa})^{1/2}(U);H)} \mathrm{d}s \right], \end{aligned}$$

where $Q^{c\kappa}$ denotes the covariance operator of the process $W^{c\kappa}(t)$. For the Hilbert-Schmidt norm we get

$$\|\psi(s)\|^2_{L_{HS}((Q^{c\kappa})^{1/2}(U);H)} = \sum_{j=\kappa+1}^{\infty} \mu_j \|\psi(s) f_j\|^2_H \leq \sum_{j=\kappa+1}^{\infty} \mu_j \|\psi(s)\|^2_{L(U;H)}.$$

Hence the truncation error can be estimated

$$\Big\| \int_0^T \psi(s) \mathrm{d}W^{c\kappa}(s) \Big\|^2_{L^2(\Omega;H)} \leq T\, \mathbb{E}\Big[\sup_{t\in[0,T]} \|\psi(s)\|^2_{L(U;H)} \Big] \left(\sum_{j=\kappa+1}^{\infty} \mu_j \right).$$

In combination with a spatial and temporal discretisation the task is now to couple the truncation parameter κ to the refinement parameter h such that the convergence error is not dominated by the error induced by the truncation of the Karhunen–Loève expansion (1.3). The following example from [45] illustrates how the decay rate of the eigenvalues of Q influences the choice of κ:

Example 1.17. *Consider the stochastic heat equation from Example 1.8. Assume that W is a Q-Wiener process with eigenvalues $\mu_k = C_\mu k^{-\alpha}$, where $C_\mu > 0$ and $\alpha > 1$, and let $F = 0$. For the spatial discretisation we choose the standard finite element method presented in Example 1.12. For the time integration we choose the backward Euler scheme*

$$X^{j+1}_{\kappa,h} = R(\Delta t A_h) \left(X^j_{\kappa,h} + \Delta t P_h F(X^j_{\kappa,h}) + P_h G(X^j_{\kappa,h}) \Delta W^{\kappa,j+1} \right), \qquad j \in \mathbb{N},$$

with $X^0_{\kappa,h} = P_h X_0$, where $R(z) = (1+z)^{-1}$ and $\Delta W^{\kappa,j+1} = W^{\kappa}(t_{j+1}) - W^{\kappa}(t_j)$. Then, we consider a noise approximation by truncating the Karhunen–Loève expansion (1.3) *at $\kappa \simeq h^{-\rho}$ for $\rho > 0$ and $\Delta t \simeq h^2$. By [45, Theorem 3.2], we get that if $\rho(\alpha - 1) = 2$ then*

there exists a constant $C > 0$ *such that* $\max_{t_k \in \Theta_K} \|X(t_k) - X^k_{\kappa,h}\|_{L^2(\Omega;H)} \leq Ch$.

Another very important issue concerning the noise approximation is the efficiency of algorithms approximating trajectories of the Q-Wiener process. As it is pointed out in [49] specific properties of eigenfunctions $(f_k, k \in \mathbb{N})$ of the covariance operator Q might be exploited to obtain fast algorithms to simulate $\Delta W^{\kappa,j}$, see also [46]:

Example 1.18. *In this example we discuss efficient methods to simulate numerical trajectories of* Q*-Wiener processes in various dimensions. Visualisations of the output of the presented algorithms are provided in Figure 1.*

- *(*Q*-Wiener process in 1D) Let* $D = (0, 1)$*. For* $j \in \mathbb{N}$ *we consider the eigenfunctions of* Q *defined by*

$$f_j(x) = \sqrt{2}\sin(\pi j x), \qquad x \in D,$$

  with eigenvalues $\mu_j = j^{-(2r+1+\varepsilon)}$ *for any* $\varepsilon > 0$*. Then following [49, Example 10.10] an efficient algorithm for approximating the increments of the* Q*-Wiener process* W *is based on a discrete sine transform (DST-1).*

- *(*Q*-Wiener process in 2D) Let* $D = (0, 1)^2$*. For* $j_1, j_2 \in \mathbb{Z}$ *let*

$$f_{j_1,j_2}(x) = \exp\big(2\pi \mathrm{i}(j_1 x_1 + j_2 x_2)\big), \qquad x \in D$$

  be eigenfunctions of Q *with corresponding eigenvalues* $\mu_{j_1,j_2} = \exp(-\alpha(j_1^2 + j_2^2))$*, where* $\alpha > 0$*. Then by [49, Example 10.12] we can simulate two independent trajectories of* W *by using only a single fast Fourier transform. Similar results are shown in [46].*

- *(*Q*-Wiener process in 3D) Let* $D = (0, 1)^3$*. For* $j_1, j_2, j_3 \in \mathbb{Z}$ *let*

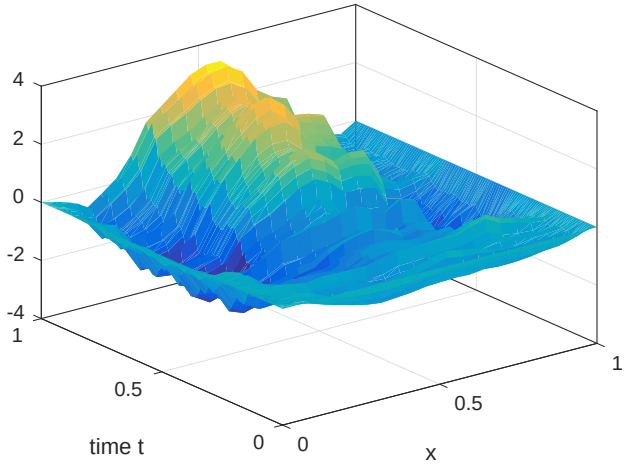
$$f_{j_1,j_2,j_3}(x) = \exp(2\pi \mathrm{i}(j_1 x_1 + j_2 x_2 + j_3 x_3)), \qquad x \in D$$

  be eigenfunctions of Q *with eigenvalues* $\mu_{j_1,j_2,j_3} = \exp(-\alpha(j_1^2 + j_2^2 + j_3^2))$ *for any* $\alpha > 0$*. By extending the results of [49, Example 10.12] to 3 dimensions we can again simulate two independent trajectories of* W *by using only a single fast Fourier transform.*

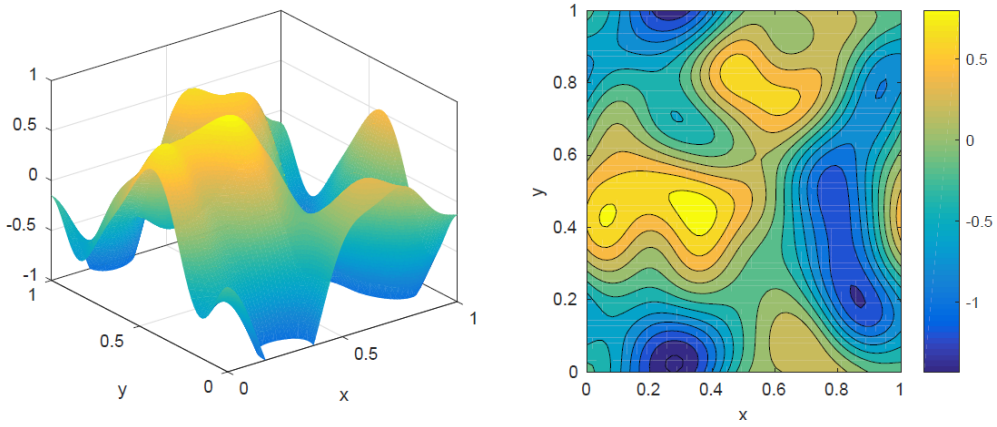
Monte Carlo methods

In many applications the statistical properties of the solution process are of high interest. For this, we want to compute the quantity of interest $\mathbb{E}[\varphi(X(T))]$, where $X(T)$ denotes the mild solution (1.5) evaluated at time T and φ is a sufficiently smooth mapping from H to another Hilbert space $(B, \langle\cdot,\cdot\rangle_B, \|\cdot\|_B)$. By definition, the expectation is represented by

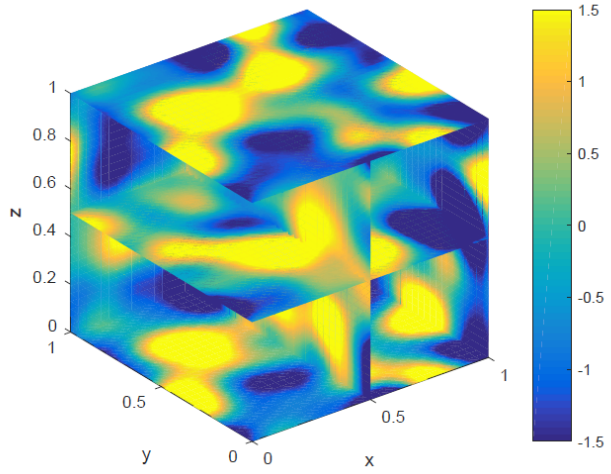
$$\mathbb{E}[\varphi(X(T))] = \int_\Omega \varphi(X(T,\omega))\mathrm{d}\mathbb{P}(\omega).$$



(a) Q-Wiener process in 1D with $r = 1$.



(b) Q-Wiener process in 2D at time $T = 1$ with $\alpha = 0.2$



(c) Q-Wiener process in 3D at time $T = 1$ with $\alpha = 0.2$

Figure 1: Visualisation of Q-Wiener process in 1D, 2D and 3D.

Solving this integral analytically is in general not possible. Hence one has to rely on suitable approximation techniques called *Monte Carlo methods*: The Monte Carlo estimator for approximating $\mathbb{E}[\varphi(X(T))]$ is defined by

$$E_M[\varphi(X(T))] = \frac{1}{M}\sum_{i=1}^{M}\varphi(X^{(i)}(T)),$$

where $(X^{(i)}(T), i = 1, \dots, M)$ are M independent realisations of the mild solution $X(T)$. For this estimator the error (typically called the *Monte Carlo* error) can be computed in the $L^2(\Omega; B)$-norm by (see [7, Lemma 4.1]):

$$\| \mathbb{E}[\varphi(X(T))] - E_M[\varphi(X(T))]\|_{L^2(\Omega;B)} = \frac{\mathrm{Var}_B[\varphi(X(T))]^{1/2}}{\sqrt{M}} \leq \frac{\|\varphi(X(T))\|_{L^2(\Omega;B)}}{\sqrt{M}},$$

where $\mathrm{Var}_B[Y] = \mathbb{E}[\|Y - \mathbb{E}[Y]\|_B^2] = \|Y\|^2_{L^2(\Omega;B)} - \|\mathbb{E}[Y]\|_B^2$ for $Y \in L^2(\Omega; B)$. In general, one has to additionally approximate the mild solution by a numerical method. Hence the overall error is given by

$$\begin{aligned}\| \mathbb{E}[\varphi(X(T))] - E_M[\varphi(X^K_{\kappa,h})]\|_{L^2(\Omega;B)} \\ \leq \| \mathbb{E}[\varphi(X(T)) - \varphi(X^K_{\kappa,h})]\|_{L^2(\Omega;B)} + \| \mathbb{E}[\varphi(X^K_{\kappa,h})] - E_M[\varphi(X^K_{\kappa,h})]\|_{L^2(\Omega;B)},\end{aligned} \tag{1.15}$$

i.e., the overall error can be estimated in terms of the weak approximation error and the Monte Carlo error. Since the Monte Carlo error converges only with $\mathcal{O}(M^{-1/2})$, one needs in order to obtain a certain accuracy $\varepsilon > 0$ in total $M = \mathcal{O}(\varepsilon^{-2})$ samples, where the simulation of a single realisation as such is already a computationally expensive task in the context of SPDE approximations. Thus, it is evident that the efficiency of standard estimators has to be enhanced in order to reduce the high computational complexity of the standard Monte Carlo method. For this reason, we discuss in Section 3 different ways how to improve the performance of Monte Carlo estimators such as importance sampling, multilevel Monte Carlo methods and efficient parallelisation of Monte Carlo algorithms.

2 Stability theory

In this section, we consider basic results on mean-square and almost sure stability theory for stochastic differential equations and their approximations. In general, stability theory investigates the effects of small changes in the initial value on the (asymptotic) qualitative behaviour of equilibrium solutions. An equilibrium solution of SDE (1.1) is a special constant solution $X_e(t) \equiv x_e \in H$ such that for all $t \geq 0$

$$AX_e(t) + F(X_e(t)) = 0 \quad \text{and} \quad G(X_e(t)) = 0.$$

Note that by this definition additive noise SDEs are excluded from our considerations here, since the second condition $G(X_e(t)) = 0$ cannot be fulfilled for this type of SDEs. Hence

alternative concepts for investigating stability in the additive noise setting have to be considered, see e.g. [17].

Subsequently, we only consider the stability properties of the underlying zero solution, since for any SDE with equilibrium solution $X_e(t)$ the zero solution is an equilibrium of the transformed SDE with solution process $Y(t) = X(t) - X_e(t)$.

Compared to the deterministic case there exists a large variety of notions of stochastic stability. In this thesis we divide the probabilistic considerations of stability into the following three types of stochastic stability in the non-asymptotic setting:

Definition 2.1. The zero solution of SPDE (1.1) is called

a) *stable in probability* if for any given $0 < \varepsilon < 1$ and $\varepsilon' > 0$, there exists a $\delta > 0$ such that if $\|X_0\|_H < \delta$ holds $\mathbb{P}$-a.s., then $\mathbb{P}[|X(t)| > \varepsilon'] < \varepsilon$ for all $t \geq 0$,

b) *almost surely stable* if for any given $\varepsilon > 0$, there exists a $\delta > 0$ such that $\|X_0\|_H < \delta$ $\mathbb{P}$-a.s. guarantees $\mathbb{P}[\|X(t)\|_H < \varepsilon \text{ for all } t \geq 0] = 1$,

c) *mean-square stable* if for any given $\varepsilon > 0$, there exists a $\delta > 0$ such that $\mathbb{E}[\|X_0\|_H^2] < \delta$ guarantees $\mathbb{E}[\|X(t)\|_H^2] < \varepsilon$ for all $t \geq 0$.

If we consider the asymptotic qualitative behaviour of the zero solution of (1.1), then we distinguish between the following concepts of asymptotic stability:

Definition 2.2. The zero solution of SPDE (1.1) is called

a) *asymptotically stable in probability* if it is stable in probability and for each $\varepsilon > 0$ there exists a $\delta > 0$ such that $\|X_0\|_H < \delta$ $\mathbb{P}$-a.s. guarantees $\lim_{t\to\infty} \mathbb{P}[\|X(t)\|_H > \varepsilon] = 0$,

b) *asymptotically almost surely stable* if it is stable in probability and there exists a $\delta > 0$ such that $\|X_0\|_H < \delta$ $\mathbb{P}$-a.s. guarantees $\mathbb{P}[\lim_{t\to\infty} \|X(t)\|_H = 0] = 1$,

c) *asymptotically mean-square stable* if it is mean-square stable and there exists a $\delta > 0$ such that $\mathbb{E}[\|X_0\|_H^2] < \delta$ guarantees that $\lim_{t\to\infty} \mathbb{E}[\|X(t)\|_H^2] = 0$.

If the zero solution of SPDE (1.1) is not stable in any of the above mentioned concepts of stochastic stability, then it is unstable in the corresponding sense. The above list of different notions of stochastic stability is by far not complete, since there exist additional concepts describing the qualitative behaviour of the zero solution of (1.1) such as *stability in the pth moment* or *stability in the large*, see e.g. [38, 47, 51] for an overview. However, the collection from above of different types of stochastic stability is complete in the sense that these characterisations cover all concepts that are treated in this thesis.

The following example of the scalar geometric Brownian motion illustrates the connections of the different concepts of stochastic stability:

Example 2.3. *Consider the geometric Brownian motion, i.e., for $\lambda, \sigma \in \mathbb{R}$ and a standard, scalar Brownian motion $(\beta(t), t \geq 0)$ let*

$$\mathrm{d}X(t) = \lambda X(t)\mathrm{d}t + \sigma X(t)\mathrm{d}\beta(t), \qquad t \geq 0,$$

with initial value $X(0) = x_0 \in \mathbb{R}$. It is well-known, see e.g. [4], that the strong solution of this SODE is given by

$$X(t) = x_0 \exp\left(\left(\lambda - \frac{\sigma^2}{2}\right) t + \sigma\beta(t)\right)$$

with second moment $\mathbb{E}[X(t)^2] = x_0^2 \exp((2\lambda + \sigma^2)t)$. This representation of the second moment immediately implies that the zero solution of the geometric Brownian motion is (asymptotically) mean-square stable if and only if $2\lambda + \sigma^2 < 0$. By computing the Lyapunov exponent (see Section 2.2 for details) of the geometric Brownian motion, we obtain asymptotically almost sure stability of the zero solution if $\lambda - \frac{\sigma^2}{2} < 0$. As a consequence we get that for all $\lambda \in \mathbb{R}$ there exists a value $\bar{\sigma} \in \mathbb{R}$ such that for all $\sigma \geq \bar{\sigma}$ the zero solution of the geometric Brownian motion is asymptotically a.s. stable, but asymptotically unstable in the mean-square sense. The effects of this specific stability setting on Monte Carlo estimators for estimating the second moment are studied in detail in Chapter 2.

Note that we required for the definition of asymptotic almost sure stability in the non-asymptotic setting only stability in probability. This is due to the fact that in many standard examples the solution process often satisfies $\mathbb{P}$-a.s. that $\lim_{t\to\infty} \|X(t)\|_H = 0$, however, the zero solution is never almost surely stable in the non-asymptotic sense. An example for this is given by the geometric Brownian motion from Example 2.3, for which we get that the probability for $X(t)$ being larger than any positive threshold at any time $t > 0$ is – independent of the choice of a non-zero intial value x_0 – strictly positive, i.e., for all $C > 0$ it holds that $\mathbb{P}[X(t) > C] > 0$.

As for analytical solutions of SPDE (1.1) we are also interested in the qualitative behaviour of equilibria of their numerical approximations, where we restrict ourselves to the numerical schemes (1.10) (Euler-Maruyama schemes) and (1.14) (Milstein schemes). Subsequently we assume that the equilibrium $(X_{h,e}^j, j \in \mathbb{N}_0)$ of these stochastic recurrences is given by the zero solution $(X_{h,e}^j = 0, j \in \mathbb{N}_0)$. In this thesis we investigate stability properties of numerical methods in the mean-square sense:

Definition 2.4. Let $(X_h^j, j \in \mathbb{N}_0)$ be the considered numerical approximation of the mild solution (1.5) for fixed time step size Δt and spatial refinement parameter h.

a) Then, the zero solution of the numerical approximation scheme given in (1.10) or (1.14), respectively, is called *mean-square stable* if for every $\varepsilon > 0$ there exists $\delta_1 > 0$ such that for all $j \in \mathbb{N}_0$, $\mathbb{E}[\|X_h^j\|_H^2] < \varepsilon$, whenever $\mathbb{E}[\|X_h^0\|_H^2] < \delta_1$.

b) It is called *asymptotically mean-square stable* if it is mean-square stable and there exists $\delta_2 > 0$ such that $\mathbb{E}[\|X_h^0\|_H^2] < \delta_2$ implies $\lim_{j\to\infty} \mathbb{E}[\|X_h^j\|_H^2] = 0$.

Furthermore, the zero solution is called *asymptotically mean-square unstable* if it is not asymptotically mean-square stable.

For finite-dimensional SODEs there has also been some progress towards an almost sure stability analysis of numerical approximation schemes, see e.g. [10] by using a discrete version of the Itô formula and see e.g. in [3] for results on discrete random dynamical

system theory. In this thesis we do not investigate a.s. stability properties of numerical approximations, but we want to emphasize that the extension of the finite-dimensional results to approximations of infinite-dimensional SDEs would be an interesting task for future work.

The main focus of this thesis in the context of stability is on the interplay of stochastic stability properties of the analytical solutions to SDE (1.1) and those of their numerical approximations. Within these investigations we mainly consider two different aspects of this interplay that can be expressed in terms of the following two questions that are taken in the spirit of [33]:

- *Under which conditions on the refinement parameters does the numerical approximation reproduce the qualitative behaviour of the analytical solution?*
- *How likely is it to detect the qualitative behaviour of the analytical solution by using numerical experiments with fixed refinement parameters?*

The first question is treated in Chapter 4, where we consider the mean-square stability properties of approximations of infinite-dimensional linear SDEs and – in particular – provide conditions that ensure the simultaneous mean-square stability of the zero solutions of the SPDE and of its approximation.

The second question is treated in Chapter 2, where we show that even in the simple case of Example 2.3 (geometric Brownian motion) numerical experiments based on Monte Carlo simulations estimating the second moment cannot reliably detect the right qualitative behaviour of the analytical solution in the mean-square sense.

In the following two subsections we describe two approaches how to identify the stability properties of the zero solution of an SDE and of its approximations. In Section 2.1 we investigate the stability properties of (approximations of) semilinear SPDEs by using results of Lyapunov's second method for SPDEs (Lyapunov functional approach) and afterwards, in Section 2.2 we consider structural approaches to identify the stability properties for linear SODE systems and their approximations.

2.1 Lyapunov functional approach

An important method for the mean-square stability analysis for semilinear (or even nonlinear) stochastic differential equations originates from Lyapunov's second method: The key idea is to find or construct a special functional satisfying certain conditions that ensure the stability of the zero solution, see [51, 38] for SODEs and [47] for SPDEs. For semilinear SPDEs the following result is due to [47, Proposition 3.1.1]:

Proposition 2.5. *Let Assumption 1.9 be fulfilled and assume that $F(0) = 0$ and $G(0) = 0$. Furthermore, let $X(0) = X_0 \in \dot{H}^1$ be deterministic. Suppose that there exists a nonnegative functional $\Lambda \in \mathcal{C}_b^2(H;\mathbb{R})$ and constants $p \geq 2, c > 0$ such that*

$$(\mathcal{L}\Lambda)(x) = \langle D\Lambda(x), Ax + F(x)\rangle_H + \frac{1}{2}\operatorname{tr}[G(x)QG^*(x)D^2\Lambda(x)] \leq -c\|x\|_H^p$$

for arbitrary $x \in \mathcal{D}(A)$. Then there exist constants $\mu > 0$ and $M \geq 1$ such that

$$\mathbb{E}[\|X(t)\|_H^p \leq M\|X_0\|_H^2 e^{-\mu t} \qquad \text{for } t \geq 0.$$

Note that the construction of functionals that satisfy the conditions of Proposition 2.5 is not a trivial task. However, for the stochastic heat equation from Example 1.8 we can find a suitable candidate for such a Lyapunov functional that ensures the asymptotic mean-square stability of the trivial solution under certain conditions:

Example 2.6. *Consider the stochastic heat equation on the domain $D = (0,1)$ with homogeneous Dirichlet boundary conditions, where $F \in L(H)$ and $G \in L(H; L(U;H))$. We choose $\Lambda(x) = \|x\|_H^2$, for which we can apply by results presented in Section 1.2 the infinite-dimensional Itô formula. Hence, we obtain*

$$(\mathcal{L}\Lambda)(u) = 2\langle u, Au + F(u)\rangle_H + \operatorname{tr}(G(u)QG^*(u)) \quad \text{for all } v \in \mathcal{D}(-A).$$

As it is shown in Chapter 4, if the condition

$$2(\|F\|_{L(H)} - \lambda_1) + \operatorname{tr}(Q)\|G\|_{L(H;L(U;H))}^2 < 0 \tag{2.1}$$

is fulfilled, then $\Lambda(x) = \|x\|_H^2$ is a Lyapunov functional and consequently, the zero solution of the stochastic heat equation is asymptotically mean-square stable.

So far we have only considered the mean-square stability properties of the analytical solution (1.5). However, the Lyapunov functional approach can be also extended to stochastic recurrences as e.g. the backward Euler scheme given in Equation (1.12). This has already been shown for approximations of finite-dimensional stochastic differential equations, see [41, 57, 62] and the references therein. These discrete approaches can be extended to approximations of SDEs in infinite dimensions as the following theorem shows:

Theorem 2.7. *Assume that there exists a non-negative, real-valued sequence $(V_j, j \in \mathbb{N}_0)$ and constants $c_1, c_2 > 0$ such that*

$$\mathbb{E}[V_0] \leq c_1 \mathbb{E}[\|X_h^0\|_H^2], \tag{2.2}$$

$$\mathbb{E}[\Delta V^j] \leq -c_2 \mathbb{E}[\|X_h^j\|_H^2], \qquad j \in \mathbb{N}_0, \tag{2.3}$$

where $\Delta V^j = V_{j+1} - V_j$. Then it follows that the zero solution of the numerical approximation scheme (1.10) *(Euler-Maruyama) or* (1.14) *(Milstein scheme), respectively, is asymptotically mean-square stable.*

Proof. This proof is based on the methods presented in [62].

First by using condition (2.3), we get that for all $j, k \in \mathbb{N}_0$ that

$$\mathbb{E}[\Delta V^{j+k}] \leq -c_2 \mathbb{E}[\|X_h^{j+k}\|_H^2],$$

and thus

$$\sum_{k=0}^{n} \mathbb{E}[\Delta V^{j+k}] = \mathbb{E}[V_{j+n+1}] - \mathbb{E}[V_j] \leq -c_2 \sum_{k=0}^{n} \mathbb{E}[\|X_h^{j+k}\|_H^2]. \tag{2.4}$$

Consequently we can deduce from (2.4) that

$$c_2 \sum_{k=0}^{n} \mathbb{E}[\|X_h^{j+k}\|_H^2] \leq \mathbb{E}[V_j] - \mathbb{E}[V_{j+n+1}] \leq \mathbb{E}[V_j],$$

where we used in the second inequality that $V_{j+n+1} \geq 0$. By taking the limit $n \to \infty$ we get that for all $j \in \mathbb{N}_0$

$$c_2 \sum_{k=0}^{\infty} \mathbb{E}[\|X_h^{j+k}\|_H^2] \leq \mathbb{E}[V_j].$$

Since (2.3) also implies that the sequence $(V_j, j \in \mathbb{N}_0)$ is decreasing, we get that

$$c_2 \sum_{k=0}^{\infty} \mathbb{E}[\|X_h^{j+k}\|_H^2] \leq \mathbb{E}[V_j] \leq \mathbb{E}[V_0] \leq c_1 \mathbb{E}[\|X_h^0\|_H^2] < \infty,$$

where we used (2.2) in the third inequality and finally the fact that by Assumption 1.10 we get that $X_h^0 \in L^2(\Omega; H)$. Since the series on the left hand side is finite and all summands $\mathbb{E}[\|X_h^{j+k}\|_H^2]$ are positive, we get that

$$\lim_{j \to \infty} \mathbb{E}[\|X_h^j\|_H^2] = 0,$$

which completes the proof. □

We now want to apply Theorem 2.7 to numerical approximations of the mild solution based on the Galerkin backward Euler scheme from Equation (1.12). For simplicity, we consider here the case $F = 0$ and $G(0) = 0$ in SPDE (1.1). As a candidate for the discrete Lyapunov function $(V_j, j \in \mathbb{N}_0)$ from Theorem 2.7 we consider $V_j = \|X_h^j\|_H^2$ for $j \in \mathbb{N}_0$. This specific choice is motivated by Example 2.6, where the analytical case is investigated also by using the squared Hilbert space-norm as candidate for the Lyapunov functional.

First, we note that $(V_j, j \in \mathbb{N}_0)$ is non-negative and that condition (2.4) from Theorem 2.7 is fulfilled for any constant $c_1 \geq 1$, since $\mathbb{E}[V_0] = \mathbb{E}[\|X_h^0\|_H^2] \leq c_1 \mathbb{E}[\|X_h^0\|_H^2]$. Thus, it remains to show that the condition from Equation (2.3) is fulfilled. For this, we consider

$$\mathbb{E}[\Delta V^j] = \mathbb{E}[V_{j+1} - V_j] = \mathbb{E}[\|X_h^{j+1}\|_H^2] - \mathbb{E}[\|X_h^j\|_H^2],$$

and by using the definition of the backward Euler-Maruyama scheme we obtain

$$\begin{aligned}
&\mathbb{E}[\|X_h^{j+1}\|_H^2] - \mathbb{E}[\|X_h^j\|_H^2] \\
&\quad = \mathbb{E}[\|R(\Delta t A_h)(X_h^j + P_h G(X_h^j)\Delta W^{j+1})\|_H^2] - \mathbb{E}[\|X_h^j\|_H^2] \\
&\quad \leq 2\|R(\Delta t A_h)\|_{L(H)}^2 \left(\mathbb{E}[\|X_h^j\|_H^2 + \|P_h G(X_h^j)\Delta W^{j+1}\|_H^2]\right) - \mathbb{E}[\|X_h^j\|_H^2] \\
&\quad \leq 2 \max_{k=1,\dots,N_h} |R(-\Delta t \lambda_{h,k})|^2 \left(\mathbb{E}[\|X_h^j\|_H^2 + \mathbb{E}[\|P_h G(X_h^j)\Delta W^{j+1}\|_H^2]\right) - \mathbb{E}[\|X_h^j\|_H^2].
\end{aligned}$$

By Itô's isometry we obtain

$$\begin{aligned}
\mathbb{E}[\|P_h G(X_h^j)\Delta W^{j+1}\|_H^2] \leq \mathbb{E}[\|G(X_h^j)\Delta W^{j+1}\|_H^2] &= \mathbb{E}\left[\left\|G(X_h^j)\int_{t_j}^{t_{j+1}} \mathrm{d}W(s)\right\|_H^2\right] \\
&= \mathbb{E}\left[\int_{t_j}^{t_{j+1}} \|G(X_h^j)\|_{L_{HS}(U_0;H)}^2 \mathrm{d}s\right] = \Delta t\, \mathbb{E}[\|G(X_h^j)\|_{L_{HS}(U_0;H)}^2] \\
&\leq \Delta t C_G^2 \mathbb{E}[\|X_h^j\|_H^2],
\end{aligned}$$

where C_G denotes the Lipschitz continuity constant from Assumption 1.9. Finally we get that

$$\mathbb{E}[\Delta V^j] \leq \left(2 \max_{k=1,\dots,N_h} |R(-\Delta t \lambda_{h,k})|^2 (1 + \Delta t C_G^2) - 1\right) \mathbb{E}[\|X_h^j\|_H^2].$$

Hence the zero solution of the numerical approximation scheme (1.12) is asymptotically mean-square stable if

$$2 \max_{k=1,\dots,N_h} |R(-\Delta t \lambda_{h,k})|^2 (1 + \Delta t C_G^2) < 1,$$

which by using the explicit representation of the rational approximation $R(z)$ yields

$$C_G^2 < -\frac{1}{2\Delta t} + \frac{2\lambda_{h,1} + \Delta t \lambda_{h,1}^2}{2}. \tag{2.5}$$

Note that the condition from Equation 2.5 is always fulfilled for a sufficiently large time step size Δt. This observation is strongly related to the well-known result that for finite-dimensional SODEs the zero solution of the backward Euler approximation scheme can be asymptotically mean-square stable although the zero solution of the SODE system is mean-square unstable. Similar results are available in the literature on the numerical analysis of deterministic differential equations, where this behaviour is often referred as *overdamping*.

In many applications the construction of the Lyapunov sequence $(V_j, j \in \mathbb{N}_0)$ is absolutely not trivial and it tends to be very sensitive with respect to the properties of the SPDE and the considered numerical approximation scheme. For this reason, we propose in Chapter 4 a structural technique to investigate the mean-square stability properties of numerical approximations of linear SPDEs. These techniques extend the results that are presented in Section 2.2.

2.2 Linear stability analysis

An important tool for the numerical analysis of deterministic ordinary differential equations (ODEs) is called linear stability analysis. For this we consider the ODE $\dot{x}(t) = f(x(t))$ with equilibrium x_e and let $x(t) = x_e + p(t)$, where $p(t)$ denotes a small perturbation with $|p(t)| \ll 1$. Then for a sufficiently smooth function f we obtain

$$\frac{\mathrm{d}x_e}{\mathrm{d}t} + \frac{\mathrm{d}p(t)}{\mathrm{d}t} = f(x_e) + Df(x_e)p(t) + \mathcal{O}(|p(t)|^2),$$

which is equivalent to

$$\frac{\mathrm{d}p(t)}{\mathrm{d}t} = Df(x_e)p(t) + \mathcal{O}(|p(t)|^2).$$

If $p(t)$ is small enough, then the first order term $Df(x_e)p(t)$ dominates and the system essentially reduces to

$$\frac{\mathrm{d}p(t)}{\mathrm{d}t} = Df(x_e)p(t). \tag{2.6}$$

It is a well-known result (see e.g. in [64, Theorem 2.3.2]) that if all eigenvalues of $Df(x_e)$ have strictly negative real parts, then the perturbation $p(t)$ vanishes as time $t \to \infty$. In contrast, the magnitude of the perturbation grows in time if at least one eigenvalue of $Df(x_e)$ has a positive real part.

By using [64, Theorem 2.3.5] one can connect the stability properties of the nonlinear ODE and the stability properties of the linearisation in the following way: If the equilibrium solution x_e is non-degenerate in the sense that all eigenvalues of $Df(x_e)$ have non-zero real parts, then the equilibrium x_e of the non-linear ODE is asymptotically stable. In reverse, if at least one eigenvalue has a positive real part, then the equilibrium is unstable. Hence the stability properties of the nonlinear ODE can be determined in the non-degenerate setting by investigating the stability properties of the linearised system (2.6).

In the next step one diagonalises (if this is possible) the system matrix $Df(x_e)$, which leads to a decoupling into d scalar ODEs of the form $x'(t) = \lambda x(t), \lambda \in \mathbb{C}$, which is typically referred as the (Dahlquist) test equation. It is clear that the zero solution of the test equation is asymptotically stable if and only if $\Re(\lambda) < 0$.

One obtains useful insights into the approximation quality of a numerical method by comparing the stability properties of the original test equation with the stability properties of the numerical scheme applied to this test equation. As an example, we consider the forward Euler (FE) scheme and the backward Euler (BE) scheme, respectively: For $n \in \mathbb{N}_0$ let

$$x_{n+1} = x_n + \Delta t \lambda x_n, \tag{FE}$$

$$x_{n+1} = x_n + \Delta t \lambda x_{n+1}. \tag{BE}$$

Then, the zero solution of the forward Euler scheme is asymptotically stable if and only if $|1 + \Delta t\lambda| < 1$ and the zero solution of the backward Euler scheme is asymptotically stable

if and only if $|1 - \Delta t\lambda| < 1$, which is a considerably less restrictive condition than for the forward Euler scheme.

Linear stochastic stability theory

For the linear stability theory of stochastic differential equations there are two substantial differences compared to the deterministic setting: First, results connecting the stability properties of nonlinear SDE systems with the corresponding mean-square stability properties of the linearised system are different, e.g., in [38] the mean-square stability analysis of the linearised system allows to deduce the stability properties of the nonlinear SODE only in probability - and not in the mean-square sense. Second, if we consider the linearisation of nonlinear stochastic ordinary differential equations we obtain a linear system of equations given by

$$\mathrm{d}X(t) = AX(t)\mathrm{d}t + \sum_{r=1}^{m} G_r X(t)\mathrm{d}\beta_r(t), \tag{2.7}$$

where $A, G_r \in \mathbb{R}^{d\times d}$. In contrast to the linearisation of an deterministc differential equation around an equilibrium solution, we now have $m+1$ matrices that have to be simultaneously diagonalisable in order to get a decoupled system of one-dimensional SODEs. However, this is in general not fulfilled as we have already seen in Example 1.16. Thus, in order to perform linear stability analysis of SODEs it is not sufficient to consider only one-dimensional SDEs such as the geometric Brownian motion. In [15, 18, 16], the authors investigate the effects of the geometry of the noise on the stability properties of the trivial solution. It turns out that there is substantial difference in the qualitative behaviour of the zero solution in cases, where the matrices G_r commute or where they do not. Hence, for the linear stochastic stability analysis, we also have to take higher dimensional SODE systems into account.

Linear mean-square stability theory

A useful mathematical tool for the mean-square stability analysis of linear SDEs is presented in [18], where the asymptotic mean-square stability can be identified by the spectral properties of the $(d^2 \times d^2)$-dimensional matrix

$$\mathcal{S} = A \otimes_K Id_d + Id_d \otimes_K A + \sum_{r=1}^{m} G_r \otimes_K G_r, \tag{2.8}$$

where $\otimes_K$ denotes the Kronecker matrix product: The zero solution of the linear SODE system (2.7) is asymptotically mean-square stable if and only if the spectral abscissa

$$\alpha(\mathcal{S}) = \max_{k=1,\dots,d^2}\{\Re(\lambda_k)|\lambda_k \text{ is an eigenvalue of } \mathcal{S}\} < 0.$$

As for the SODE system itself one can also derive a matrix-eigenvalue problem that identifies the asymptotic mean-square stability properties of its numerical approximations. By applying a numerical one-step time integration method (such as θ-Maruyama methods of θ-Milstein methods) to the linear SODE system (2.7) we obtain a stochastic recurrence equation of the following type:

$$X_{i+1} = \mathfrak{A}_i X_i, \qquad i \in \mathbb{N}_0, \tag{2.9}$$

where $(\mathfrak{A}_i, i \in \mathbb{N}_0)$ is a sequence of independent and identically distributed (iid for short) random matrices. By employing the vectorisation operator to $X_i X_i^T$ we obtain the relation

$$\mathbb{E}[Y_{i+1}] = \mathbb{E}[\mathfrak{A}_i \otimes_K \mathfrak{A}_i]\, \mathbb{E}[Y_i] = S\, \mathbb{E}[Y_i],$$

where $Y_i = \mathrm{vec}(X_i X_i^T) \in \mathbb{R}^{d^2}$ and $S = \mathbb{E}[\mathfrak{A}_1 \otimes_K \mathfrak{A}_1] \in \mathbb{R}^{d^2 \times d^2}$. Then by [18, Lemma 3.4] the zero solution of the system of linear recurrence equations (2.9) is asymptotically mean-square stable if and only if the spectral radius satisfies

$$\rho(S) = \max_{k=1,\dots,d^2} \{|\lambda_k| \mid \lambda_k \text{ is an eigenvalue of } S\} < 1.$$

Explicit representation formulae of the matrix S are computed for θ-Maruyama methods in [18, Theorem 3.7], and for θ-Milstein methods applied to SODEs with either commutative noise in [18, Theorem 3.9] or with non-commutative noise in [18, Theorem 3.10], where in the latter case the iterated stochastic integrals are computed by considering a truncation of the corresponding Lévy areas that preserves the strong convergence order 1 of the θ-Milstein scheme, see e.g. [23, 40, 69]. The spectral radius approach has also been extended to other numerical schemes such as e.g. stochastic Runge-Kutta schemes, see [31].

This eigenvalue approach turns out to be very effective, since this leads to a comparative analysis of numerical methods applied to the SODE (2.7). However, the complexity of computing the eigenvalues of S analytically or also numerically increases heavily if we increase the dimension d of the linear problem. This is the case for e.g. semidiscretised SPDEs, for which $d \approx N_h$ might become arbitrarily large. For this reason, we develop in Chapter 4 a systematic approach to detect the mean-square stability properties of numerical approximations applied to linear SPDEs based on properties of the involved operators of SPDE (1.1) as well as of the considered numerical approximation scheme.

Linear almost sure stability analysis

Investigating stochastic stability in the almost sure sense turns out to be significantly more complicated than the stability analysis in the mean-square sense. A large class of the techniques to perform almost sure stability analysis has its origins in the random dynamical system theory, see the classical monograph [3] for a detailed overview. For further discussions on the random dynamical system approach, we have to restrict ourselves to analysing

linear SODE systems (2.7), for which the following non-degeneracy (or also called hypoellipticity) condition is fulfilled:

Assumption 2.8. *Assume that for all $x, y \in \mathbb{R}^d$ there exists a constant $C > 0$ such that*

$$\sum_{r=1}^{m} \langle G_r x, y\rangle^2 \geq C|x||y|.$$

Note that there are several possibilities how this assumption can be verified, see [38, Appendix B]. For instance, the condition that $\dim L(G_1 x, \ldots, G_m x) = d$ for all $x \in \mathbb{R}^d, x \neq 0$, where L denotes the linear hull spanned by the vectors $G_i x$, is equivalent to Assumption 2.8.

The asymptotic almost sure stability properties of the zero solution of (2.7) are strongly related to the Lyapunov exponents $\Lambda_d \leq \cdots \leq \Lambda_1$ of the system (2.7): By the multiplicative ergodic theorem (see e.g. [3, Theorem 3.4.11]), there exists a splitting of the initial value set $\mathbb{R}^d$ into random subspaces $E_i(\omega)$, i.e.,

$$\mathbb{R}^d = \bigoplus_{i=1}^{d} E_i(\omega),$$

such that for almost all $\omega \in \Omega$ and $i = 1, \ldots, d$, it holds that

$$0 \neq x_0 \in E_i(\omega) \quad \Leftrightarrow \quad \lim_{t\to\infty} \frac{1}{t} \log |X(t; x_0)| = \Lambda_i, \qquad \mathbb{P}\text{-a.s.}$$

It has been shown in [3, Theorem 6.2.16] that if the initial value x_0 is deterministic and if Assumption 2.8 is fulfilled that

$$\Lambda_1 = \lim_{t\to\infty} \frac{1}{t} \log |X(t; x_0)| \qquad \mathbb{P}\text{-a.s.},$$

i.e., the asymptotic behaviour of the solution process $X(t)$ in the almost sure sense can then be determined by examining the upper Lyapunov exponent Λ_1. By [38, Theorem 6.11] we get under Assumption 2.8 that the zero solution of (2.8) is asymptotically almost surely stable if $\Lambda_1 < 0$. In reverse the same theorem proves that the zero solution is asymptotically unstable in the almost sure sense if $\Lambda_1 > 0$.

For computing Λ_1 we can proceed as it is presented in [38, Section 6.7]: First we consider the projection of the solution process $X(t)$ onto the $(d-1)$-dimensional sphere $S^{d-1} = \{v \in \mathbb{R}^d : \|v\|_{\mathbb{R}^d} = 1\}$, i.e., we consider the process

$$S(t) = \frac{X(t)}{|X(t)|},$$

together with

$$\rho(t) = \log |X(t)|.$$

This can be seen as the separation of the process into the radial and the angular parts. The radial part $\rho(t)$ is then given as the solution of the scalar SODE

$$\mathrm{d}\rho(t) = \mathcal{K}(S(t))\mathrm{d}t + \sum_{r=1}^{m} \langle G_r S(t), S(t)\rangle \, \mathrm{d}\beta_r(t),$$

where

$$\mathcal{K}(\lambda) = \langle F\lambda, \lambda\rangle + \frac{1}{2}\operatorname{tr} D(\lambda) - \langle D(\lambda)\lambda, \lambda\rangle$$

for $D(\lambda) = \sum_{r=1}^{m} G_r \lambda\lambda^T G_r \in \mathbb{R}^{d\times d}$. Then it holds that

$$\Lambda_1 = \int_{S^{d-1}} \mathcal{K}(\lambda)\nu(\mathrm{d}\lambda), \tag{2.10}$$

where $\nu(\mathrm{d}\lambda)$ denotes the unique normalized invariant measure of the process $S(t)$, which exists since Assumption 2.8 is fulfilled.

Unfortunately finding an explicit representation of Λ_1 has only been possible for rather low dimensional linear systems of SODEs (2.7), where the dimension is either $d = 1$ (geometric Brownian motion) or $d = 2$, see e.g. [37, 34, 3]. The direct computation of the top Lyapunov exponent by using the representation (2.10) faces severe difficulties for higher dimensions ($d \geq 3$). Another attempt that has been proposed in [65] is based on numerically estimating the upper Lyapunov exponent on a large time horizon, which is in general also computationally expensive.

Instead of computing the explicit representation of Λ_1, we provide in this theses estimates of Λ_1 for a spatially discretised stochastic heat equation that can be used to determine the almost sure stability properties of the underlying zero solution. For this the main idea in Chapter 2 is to examine the spectral properties of the kernel $\mathcal{K}(\lambda), \lambda \in S^{d-1}$, of the integral representation in Equation (2.10).

3 Efficiency of numerical methods

In this section we investigate the efficiency of numerical algorithms for approximating the quantity of interest $\mathbb{E}[\varphi(X(T))]$, where $X(T)$ denotes the mild solution of SPDE (1.1) evaluated at time T. In this thesis, *efficiency* is understood as a measure for the computational cost to obtain a certain accuracy.

As we have already seen in Section 1.3 standard Monte Carlo techniques are computationally expensive, since a large number of samples has to be simulated in order to reduce the approximation error to a certain accuracy. For this reason we discuss here three different techniques how to improve the performance of standard Monte Carlo estimators. We first consider variance reduction techniques based on importance sampling in Section 3.1. Then, in Section 3.2 we consider multilevel Monte Carlo methods and finally, we discuss in Section 3.3 how fully parallel algorithms can be used for enhancing the performance of Monte Carlo methods.

3.1 Importance Sampling

Importance sampling is a frequently used approach to reduce the variance of a considered random variable Y for various applications of Monte Carlo simulations. The basic idea is to transform the underlying probability measure $\mathbb{P}$ to another $\widetilde{\mathbb{P}}$, from which we sample the realisations of a transformed version of Y in order to weight *important* outcomes stronger.

In the literature, importance sampling has already been applied to SDEs in finite dimensions, see e.g. [39, 55]: Consider the following finite-dimensional SODE system

$$\mathrm{d}X(t) = f(X(t))\mathrm{d}t + \sum_{r=1}^{m} g_r(X(t))\mathrm{d}\beta_r(t), \qquad \text{for } t \in [0,T], \tag{3.1}$$

and the related SODE system with modified drift

$$\mathrm{d}\widetilde{X}(t) = f(\widetilde{X}(t)) - \sum_{r=1}^{m} g_r(\widetilde{X}(t))d_r(t,\widetilde{X}(t))\mathrm{d}t + \sum_{r=1}^{m} g_r(\widetilde{X}(t))\mathrm{d}\beta_r(t), \qquad \text{for } t \in [0,T],$$

with initial values $X(0) = \widetilde{X}(0) = x_0 \in \mathbb{R}^d$. Here the functions $d_j : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ are chosen such that the *density process* $(\Theta(t), t \in [0,T])$ given as the strong solution of the linear SODE

$$\mathrm{d}\Theta(t) = \sum_{r=1}^{m} d_r(t,\widetilde{X}(t))\Theta(t)\mathrm{d}\beta_r(t), \qquad \text{for } t \in [0,T], \tag{3.2}$$

with initial value $\Theta(0) = 1$, is a martingale. By the well-known Girsanov theorem, see e.g. [56, 60], there exists a probability measure $\widetilde{\mathbb{P}}$ such that the stochastic processes

$$\widetilde{\beta}_r(t) = -\int_0^t d_r(s,\widetilde{X}(s))\mathrm{d}s + \beta_r(t), \qquad \text{for } r = 1,\dots,m, \text{and } t \in [0,T],$$

are Brownian motions under the probability measure $\widetilde{\mathbb{P}}$. Thus, we get the identity

$$\mathbb{E}_{\mathbb{P}}[\varphi(X(T))] = \mathbb{E}_{\widetilde{\mathbb{P}}}[\varphi(\widetilde{X}(T))] = \mathbb{E}_{\mathbb{P}}[\varphi(\widetilde{X}(T))\Theta(T)],$$

where $\mathbb{E}_{\mathbb{Q}}[Y]$ denotes the expectation of a random variable with respect to the probability measure $\mathbb{Q}$. This identity leads now to the definition of the importance sampling Monte Carlo estimator (see e.g. [39, 55])

$$E_M[\varphi(\widetilde{X}(T))\Theta(T)] = \frac{1}{M}\sum_{k=1}^{M} \varphi(\widetilde{X}^{(i)}(T))\Theta^{(i)}(T),$$

where $\varphi(\widetilde{X}^{(i)}(T))\Theta^{(i)}(T)$ are M independent realisations of $\varphi(\widetilde{X}(T))\Theta(T)$. By the same arguments as for the standard Monte Carlo estimator in Section 1.3, the error of the impor-

tance sampling Monte Carlo estimator in the $L^2(\Omega;\mathbb{R})$-norm is given by

$$\|E_M[\varphi(\widetilde{X}(T))\Theta(T)] - \mathbb{E}[\varphi(X(T))]\|_{L^2(\Omega;\mathbb{R})} = \frac{\mathrm{Var}[\varphi(\widetilde{X}(T))\Theta(T)]^{1/2}}{\sqrt{M}}.$$

This motivates the use of importance sampling as a variance reduction technique: The goal is to choose admissible functions $d_j, j = 1, \dots, m$, such that the considered transformation leads to a variance reduction for the sampled random variables, i.e.,

$$\mathrm{Var}[\varphi(\widetilde{X}(T))\Theta(T)] \ll \mathrm{Var}[\varphi(X(T))]. \tag{3.3}$$

Remark 3.1. In general one has to additionally consider approximations of the random variables $\widetilde{X}^{(i)}(T)$ and $\Theta^{(i)}(T)$ since they are only known explicitly in very few cases. As a consequence, the overall error decomposes – as already seen for the standard Monte Carlo estimator in Section 1.3 – into the weak approximation error and the variance of the numerical approximation of the considered random variables. However, under the assumption (see also [55]) that the variance of the original random variable is close to the variance of its numerical approximation, variance reduction can be also obtained for the numerical approximation if (3.3) is fulfilled.

Remark 3.2. Note that numerical methods for approximating the density process $\Theta(t), t \in [0,T]$, given in Equation (3.2) have to be chosen carefully, since instabilities for coarse time steps might appear for SODEs without drift components as it has been investigated in [52].

Optimal measure transformations in the sense that the variance of the random variable $\varphi(\widetilde{X}(T))\Theta(T)$ is zero (and for this reason also the Monte Carlo error vanishes) have already been stated in the literature. In [55], it is shown for finite-dimensional SODE systems that choosing

$$d_r(t,\widetilde{X}(t)) = -\frac{1}{u(t,\widetilde{X}(t))}\left\langle g_r(t,\widetilde{X}(t)), \frac{\partial u}{\partial x}(t,\widetilde{X}(t))\right\rangle, \qquad \text{for } r = 1,\dots,m,$$

implies that $\varphi(\widetilde{X}(T))\Theta(T)$ is deterministic. However, the function u denotes the solution of the corresponding Kolmogorov backward equation, which essentially requires the full knowledge of an analytical representation of a function $u(x) = \mathbb{E}[\varphi(X(T;x)]$, where $X(t;x)$ denotes the solution of SODE (3.1) at time t with initial value $X(0) = x$. Consequently the optimal measure transformation cannot be implemented directly in numerical approximations and also finding a suitable approximation $\hat{u}$ of the solution u of the Kolmogorov backward equation is well-known to be a computationally demanding task as it is shown in [53]. For this reason we emphasize that a computationally relevant measure transformation should balance the following two properties: First, the variance should be reduced as it is stated in Equation (3.3) and second, the measure transformation should be easily computable. There are also extensions of the above optimality results to SDEs in infinite dimensions, see [54], where a combined approach of importance sampling together with a control variate approach is proposed for strong solutions of (1.1) and the specific, finite-dimensional

choice $U = \mathbb{R}^m$.

In this thesis we focus in Chapters 2 and 3 on importance sampling techniques for SPDEs. We discuss two different approaches how importance sampling can be used to reduce the variance of the simulated random variables in numerical experiments. First, we consider in Chapter 2 a method-of-lines approach, where we discuss an importance sampling technique applied to a semidiscretised version of the stochastic heat equation (by using finite differences). We are able to improve the performance of Monte Carlo estimators approximating the second moment of the solution process in cases, where the zero solution of the considered finite-dimensional SODE system is asymptotically almost surely stable, but mean-square unstable.

As second contribution to importance sampling techniques for SPDE (1.1), we extend the results of finite-dimensional importance sampling techniques to the infinite-dimensional framework. We propose a large class of infinite-dimensional measure transformations and discuss their numerical approximations in detail. As a benchmark for the design of a suitable measure transformation we develop an optimal measure transformation for SPDE (1.1) driven by a Q-Wiener process that attains values in an infinite-dimensional Hilbert space U.

3.2 Multilevel Monte Carlo methods

The aim of this section is to review different results on multilevel Monte Carlo (MLMC) techniques for approximating the expectation $\mathbb{E}[Y]$ of a random variable $Y \in L^2(\Omega; B)$. The MLMC method is well-known to be a computationally efficient method that is applicable for a large range of problems, see e.g. [28] for an extensive overview. For this reason the MLMC estimator is discussed in each of the remaining chapters from various view points, where in particular stability issues are discussed in Chapters 2 and 4 and we propose efficient numerical methods that can be coupled with MLMC techniques in Chapters 3 and 5. In this thesis we distinguish between two types of MLMC methods, i.e., MLMC methods only in time (for finite-dimensional SODE systems) and MLMC methods in space and time for approximating statistics of the mild solution process (1.5) of SPDE (1.1).

Before we discuss these two types of MLMC methods separately, we introduce the concept of multilevel Monte Carlo techniques in an abstract framework: For this we consider a sequence $(Y_\ell, \ell \in \mathbb{N}_0)$ of numerical approximations of a random variable $Y \in L^2(\Omega; B)$ on different refinement levels $\ell \in \mathbb{N}_0$ with increasing accuracy and also with increasing computational cost. These refinement levels can be understood either with respect to the time discretisation for SODEs, or – in the case of SPDEs – with respect to a combination of a discretisation in space and time. In both cases we associate each refinement level ℓ with a finite-dimensional space $\mathcal{V}_\ell$ such that $Y_\ell \in \mathcal{V}_\ell$. In many applications of the MLMC method the sequence $(\mathcal{V}_\ell, \ell \in \mathbb{N}_0)$ is chosen to be nested in the sense that $\mathcal{V}_{\ell-1} \subset \mathcal{V}_\ell$ for $\ell \in \mathbb{N}_0$. Although it is shown in [32] that a nested, geometric choice of these approximation spaces is under specific conditions not optimal, the benefits of choosing an optimal non-geometric hierarchy are negligible in our setting as it is indicated in [28]. For this reason, we subsequently assume that the sequence $(\mathcal{V}_\ell, \ell \in \mathbb{N}_0)$ is nested.

The basic idea of the MLMC estimator is based on the following observation: For $L \in \mathbb{N}_0$, we get that

$$\mathbb{E}[Y_L] = \mathbb{E}\left[\sum_{\ell=0}^{L} Y_\ell - Y_{\ell-1}\right] = \sum_{\ell=0}^{L} \mathbb{E}\left[Y_\ell - Y_{\ell-1}\right],$$

where we define $Y_{-1} = 0$ for notational convenience. By approximating each of the expectations by a standard Monte Carlo estimator (as defined in Section 1.3), we define for $L \in \mathbb{N}_0$ the multilevel Monte Carlo estimator by

$$E^L[Y] = \sum_{\ell=0}^{L} E_{M_\ell}[Y_\ell - Y_{\ell-1}],$$

where $M_\ell \in \mathbb{N}$ denotes the number of independent realisations of the random variables $Y_\ell - Y_{\ell-1}$ for the Monte Carlo estimator on level ℓ.

For providing a bound on the computational complexity of the Monte Carlo estimator the following assumption is needed:

Assumption 3.3. *Assume that there exist constants $\alpha, \beta, \theta, c > 0$ such that $\alpha \geq \frac{1}{2}\min(\beta, \theta)$ and for all $\ell \geq 0$*

a) $\|\mathbb{E}[Y_\ell - Y]\|_B \leq c2^{-\alpha\ell}$, *b)* $\mathrm{Var}_B[Y_\ell - Y_{\ell-1}] \leq c2^{-\beta\ell}$, *c)* $C[Y_\ell - Y_{\ell-1}] \leq c2^{\theta\ell}$,

where $C[\hat{X}]$ denotes the computational cost to simulate the random variable $\hat{X}$.

Under Assumption 3.3 we get by [28] that there exists a constant $\bar{c} > 0$ such that for any sufficiently small $\varepsilon > 0$ there exist values $L \in \mathbb{N}_0$ and $M_\ell \in \mathbb{N}, \ell = 0, \ldots, L$, such that

$$\mathbb{E}[\|E^L[Y] - \mathbb{E}[Y]\|_B^2] < \varepsilon^2$$

with a computational complexity – here denoted by $C[E^L[Y]]$ – with expected bound

$$\mathbb{E}[C[E^L[Y]]] \leq \begin{cases} \bar{c}\varepsilon^{-2}, & \beta > \theta, \\ \bar{c}\varepsilon^{-2}(\log\varepsilon)^2, & \beta = \theta, \\ \bar{c}\varepsilon^{-2-(\theta-\beta)/\alpha}, & \beta < \theta. \end{cases} \tag{3.4}$$

In Assumption 3.3 the parameter β denotes the decay rate of the variance of the considered random variables for the Monte Carlo estimator on level ℓ and the parameter θ measures the increase of computational work needed for simulating the realisations $(Y_\ell - Y_{\ell-1})$. For instance, it holds that $\beta = \theta$ for numerically approximating trajectories of finite-dimensional SODE systems by using standard numerical integrators such as the Euler-Maruyama scheme, see e.g. [29].

Multilevel Monte Carlo methods in time

In this part we briefly review basic results on the multilevel Monte Carlo methods applied to the d-dimensional SODE

$$\begin{aligned} \mathrm{d}X(t) &= f(X(t))\mathrm{d}t + g(X(t))\mathrm{d}\beta(t), \qquad t \in [0,T] \\ X(0) &= x_0 \in \mathbb{R}^d, \end{aligned} \tag{3.5}$$

where $(\beta(t), t \in [0,T])$ denotes an m-dimensional Brownian motion and f and g fulfil Assumption 1.9. Furthermore we consider the case that the strong solution is approximated by a numerical method with weak convergence order $\eta = 1$. For the standard Monte Carlo estimator $E_M[\varphi(X(T))]$, where $\varphi : \mathbb{R}^d \to \mathbb{R}$ is a sufficiently smooth function, we discussed in Section 1.3 that the overall mean-square error can be estimated by Equation (1.15) in terms of the Monte Carlo error and of the systematic error (induced by the weak approximation error of the numerical method). Hence, by choosing the (explicit or implicit) Euler-Maruyama approximation for the time integration, the overall (*root mean-square*) error satisfies

$$\|\mathbb{E}[\varphi(X(T))] - E_M[\varphi(X_K)]\|_{L^2(\Omega;\mathbb{R})} = \mathcal{O}(M^{-1/2} + \Delta t),$$

where we used results from [66, 39] that provide the weak convergence order 1 of the Euler-Maruyama scheme for a sufficiently smooth function φ. Thus, in order to obtain an overall error of order ε, the number of samples has to be chosen to be $M = \mathcal{O}(\varepsilon^{-2})$ and the time step size has to be chosen such that $\Delta t = \mathcal{O}(\varepsilon)$. Hence the overall computational complexity to obtain an accuracy of order ε is given by $\mathcal{O}(\varepsilon^{-3})$.

Historically the development of multilevel Monte Carlo algorithms for finite-dimensional SODEs started in [36] with the analysis of a two-level Monte Carlo estimator

$$E^1[\varphi(X(T))] = E_{M_0}[\varphi(X_0^{K_0})] + E_{M_1}[\varphi(X_1^{K_1}) - \varphi(X_0^{K_0})],$$

where for the two refinement levels $\ell = 0, 1$, we denote by $X_\ell^{K_\ell}$ a numerical approximation of $X(T)$ by using the Euler-Maruyama scheme based on partitions Θ_{K_ℓ} of the time interval $[0,T]$ with K_ℓ time steps. A complexity analysis for this method – originally called *statistical Romberg extrapolation* – showed a reduction of the computational cost to $\mathcal{O}(\varepsilon^{-5/2})$.

The idea of the two-level Monte Carlo methods is extended in [29] to multiple levels, which leads to the definition of the MLMC estimator in time

$$E^L[\varphi(X(T))] = E_{M_0}[\varphi(X_0^{K_0})] + \sum_{\ell=1}^{L} E_{M_\ell}[\varphi(X_\ell^{K_\ell}) - \varphi(X_{\ell-1}^{K_{\ell-1}})].$$

Note that by the same argumentation as in [28, Chapter 5] there exists a constant $C > 0$ such that

$$\mathrm{Var}[\varphi(X_\ell^{K_\ell}) - \varphi(X_{\ell-1}^{K_{\ell-1}})] \leq C\left(\mathbb{E}[|\varphi(X(T)) - \varphi(X_\ell^{K_\ell})|^2] + \mathbb{E}[|\varphi(X(T)) - \varphi(X_{\ell-1}^{K_{\ell-1}})|^2]\right).$$

Thus, for a geometrical sequence $(\Theta_{K_\ell}, \ell = 1, \dots, L)$ of partitions of the time interval, where for each Θ_{K_ℓ} the corresponding time step size is denoted by Δt_ℓ, we get that

$$\mathrm{Var}[\varphi(X_\ell^{K_\ell}) - \varphi(X_{\ell-1}^{K_{\ell-1}})] = \mathcal{O}(\Delta t_\ell).$$

Now, by choosing $\Delta t_\ell = 2^{-\ell}\Delta t_0$, Assumption 3.3 is fulfilled with $\alpha = 1, \beta = 1$ and $\theta = 1$. Hence, by (3.4) we get that the expected cost of the MLMC estimator to obtain an $L^2(\Omega; \mathbb{R})$-error of $\mathcal{O}(\varepsilon)$ is of order $\varepsilon^{-2}(\log \varepsilon)^2$. This improvement of the complexity is nearly optimal for estimating $\mathbb{E}[\varphi(X(T))]$, since it is shown in [20] that the lower bound on the complexity is given by $\mathcal{O}(\varepsilon^{-2})$.

Although MLMC methods are widely applicable, there are also cases, where they do not improve the approximation quality significantly compared to standard Monte Carlo methods. In this thesis we illustrate in Chapter 2 the failure of the MLMC method in time if the outcome on any refinement $\ell \in \mathbb{N}_0$ heavily depends on rare events. This specific setting is obtained e.g. by approximating the second moment of the solution process of linear SDEs, for which the zero solution is asymptotically almost surely stable, but mean-square unstable. As a remedy we propose an importance sampling MLMC technique, which forces these rare events to happen more frequently.

Multilevel Monte Carlo methods in space and time

In this section we approximate $\mathbb{E}[\varphi(X(T))]$ by using MLMC methods, where $X(T)$ denotes the mild solution of SPDE (1.1). For the spatial discretisation of the mild solution (1.5), we consider Galerkin methods based on a nested sequence $(V_{h_\ell}, \ell \in \mathbb{N}_0)$ of finite-dimensional spaces satisfying $V_{h_\ell} \subset \dot{H}^1 \subset H$ with dimension $\dim(V_{h_\ell}) = N_{h_\ell}$. Furthermore for $t_k^\ell \in \Theta_{K_\ell}$, we denote by $X_{h_\ell}^k \in V_{h_\ell}$ the refinement level ℓ-approximation of $X(t_k^\ell)$ based on a time integration method scheme such as given in Equation (1.10) or (1.14), where Θ_{K_ℓ} is the equidistant partition of the time interval $[0, T]$ with time step size $\Delta t_\ell = T/K_\ell$. Then, the MLMC estimator in space and time is defined by

$$E^L[\varphi(X(T))] = E_{M_0}[\varphi(X_{h_0}^{K_0})] + \sum_{\ell=1}^{L} E_{M_\ell}\left[\varphi(X_{h_\ell}^{K_\ell}) - \varphi(X_{h_{\ell-1}}^{K_{\ell-1}})\right].$$

Following [7, 8] we distinguish for the analysis of the mean-square error between numerical methods, for which results on the weak convergence order are available (e.g., for SPDEs with additive noise) and for approximation schemes for which only strong convergence order results have been shown.

We start with numerical approximation schemes, for which only strong convergence order results are available. Then for any $\epsilon > 0$, the choice

$$\Delta t_\ell \simeq h_\ell^2, \qquad M_0 \simeq h_L^{-2\gamma}, \quad M_\ell \simeq h_L^{-2\gamma} h_\ell^{2\gamma} \ell^{1+\epsilon}$$

guarantees the existence of a constant $C > 0$ such that (see [44, Corollary 2])

$$\| \mathbb{E}[\varphi(X(T))] - E^L[\varphi(X(T))] \|_{L^2(\Omega;B)} \leq C h_L^{\gamma}.$$

If weak convergence rates η are available and $\eta = 2\gamma$, then by choosing

$$\Delta t_\ell \simeq h_\ell^2, \qquad M_0 \simeq h_L^{-4\gamma}, \quad M_\ell \simeq h_L^{-4\gamma} h_\ell^{2\gamma} \ell^{1+\epsilon}$$

the convergence order of the MLMC estimator can be improved, since there exists a constant $C > 0$ such that (see [44, Corollary 3])

$$\| \mathbb{E}[\varphi(X(T))] - E^L[\varphi(X(T))] \|_{L^2(\Omega;B)} \leq C h_L^{2\gamma}.$$

Now we consider again the computational work needed for the $L^2(\Omega; B)$-error of the multilevel Monte Carlo estimator to be smaller than a given accuracy ε. For a geometric sequence of finite-dimensional approximation spaces based on spatial refinement parameters $h_\ell \simeq 2^{-\ell}$ the computational work can be essentially reduced by using the MLMC technique as it can be seen in Table 1 taken from [44] for the ideal convergence parameter $\gamma = 1$.

MC	MLMC with strong conv.	MLMC with weak conv.
$\varepsilon^{-(d/2+3)}$	$\varepsilon^{(-d+2)} \lvert \log_2 \varepsilon \rvert$	$\varepsilon^{-(d/2+2)} \lvert \log_2 \varepsilon \rvert$

Table 1: Order of computational work to achieve an $L^2(\Omega; B)$-error of $\mathcal{O}(\varepsilon)$ of different Monte Carlo-type approximations.

This result shows the importance of investigating weak convergence rates for multiplicative noise SPDEs, since such results would improve estimates on the computational complexity of MLMC methods in space and time significantly.

An important question that arises for the implementation of MLMC methods for SPDEs is how to choose the different levels in order to guarantee that any numerical method approximating the mild solution the SPDE (1.1) is stable enough on each refinement level. Especially the choice of the coarsest levels is of great importance, since many explicit time integrators applied to parabolic SPDEs face a severe time step size restriction and might be very inefficient as it is shown in e.g. [1]. Even worse, choosing the coarsest refinement levels such that the numerical method is not stable enough (in an appropriate stochastic sense), standard MLMC techniques do not provide reasonable approximations of the quantity of interest.

For this reason, we perform in Chapter 4 a detailed mean-square stability analysis of SPDE approximations. We prove that if Equation (2.1) from Example 2.6 is fulfilled then the zero solutions of the SPDE and of its approximation by the backward Euler–Maruyama scheme (1.12) are both asymptotically mean-square stable without any restrictions on the refinement parameters h and Δt. Similar results are also shown for the backward Milstein scheme from Equation (1.14). This shows that these methods are sufficiently stable in the

mean-square sense for applying MLMC methods in space and time to this type of SPDEs.

3.3 Parallel Scientific Computing

Recall that a Monte Carlo estimator consists of a linear combination of independent samples of the considered random variable Y. Due to this specific structure, Monte Carlo methods are well-known to be easily parallelisable. The idea of distributing the simulation of these independent realisations to different processors generally leads to optimal parallelisation properties of Monte Carlo-based algorithms.

Nevertheless, there might also occur some problems of this natural parallelisation approach that limit the efficiency of parallel Monte Carlo estimation, since the parallel generation of pseudo-random numbers is not trivial task. By an inappropriate call of standard pseudo-random number generators in parallel, an additional bias can deteriorate the performance of the Monte Carlo estimator and specific treatments of parallel random generation have to be used, see e.g. [12]. However, there are some standard libraries and implemented functions in C++ and Matlab that can handle these parallel computations and, subsequently, we always assume that the numerical experiments are based on such appropriate routines.

As long as the number of realisations of Y is larger than the number of available cores, the parallelisation of Monte Carlo simulations performs very well due to the fact that the algorithms require only a small amount of communication between the used cores. Since nowadays the tendency in computer science is to constantly increase the number of cores in supercomputers, the parallelised Monte Carlo method can be inefficient if the number of cores becomes smaller than the required sample size. In the context of numerical approximations of SDEs one has then to consider algorithms that are capable to compute also pathwise approximations in parallel.

Such an approach can be found e.g. in [9], where the authors proposed to use a (parallelisable) algebraic multigrid method for solving the linear system corresponding to a finite element discretisation of an elliptic PDE with random coefficients. This can be also exploited for time-dependent problems, for which one can use a parallelisable method to solve the linear system connected to the problem in space for each time step. In terms of optimal parallelisation this again might not be sufficient if the number of available cores is larger than the number of those that are needed.

In the deterministic framework there are approaches how to compute also time approximations of differential equations completely in parallel. For instance in [25, 24], the authors do not apply the backward Euler scheme to linear determinstc differential equations sequentially, but use an equivalent formulation of the approximation schemes as a large linear system. This enables the use of multigrid methods that can be applied to solve the whole space-time formulation at once. In [25], such an approach has been used to solve a linear ODE in parallel and in [24] a space-time multigrid method is proposed that has impressive strong and weak scaling properties when it is applied to parabolic PDEs. Here strong scaling means that the computational time reduction is proportional to the number of cores used. In contrast optimal weak scaling is achieved whenever increasing the dimension of

the problem in the same way as increasing the number of cores leaves the computation time constant.

In Chapter 5 of this thesis we combine the space-time multigrid techniques from [24] with multilevel Monte Carlo methods for parabolic SPDEs driven by additive Wiener noise. The resulting algorithm can be fully parallelised, i.e., it can be performed in parallel in space, time and probability. Extensive numerical experiments show the convergence and parallelisation properties of the proposed approach.

4 Information on the included articles

This thesis consists of an introductory essay in Chapter 1, which is followed by the following 4 included articles:

- Chapter 2 was published as "*An importance sampling technique in Monte Carlo methods for SDEs with a.s. stable and mean-square unstable equilibrium*" by *M. Ableidinger, E. Buckwar* and *A. Thalhammer* in Journal of Computational and Applied Mathematics, Vol. 316 (2017), pp. 3-14.

- Chapter 3 is a preprint version of "*Importance sampling techniques for stochastic partial differential equations*" by *E. Buckwar* and *A. Thalhammer*, which will be submitted to *Mathematics of Computation* in summer 2017. A previous version of this article has been published as *DK-Report No. 2017-01*.

- Chapter 4 is a preprint version of "*Mean-square stability analysis of approximations of stochastic differential equations in infinite dimensions*" by *A. Lang, A. Petersson* and *A. Thalhammer*, which was submitted to BIT Numerical Mathematics in February 2017. The preprint is available on *arXiv* with index *arXiv:1702.07700[math.NA]*

- Chapter 5 is a preprint version of "*Combining space-time multigrid techniques with multilevel Monte Carlo methods for SDEs*" by M. Neumüller and A. Thalhammer, which will be submitted to *SIAM Journal on Scientific Computing* in summer 2017. A previous version was submitted for the proceedings of the *International Conference on Domain Decomposition 2017 (DD24)* in May 2017 and has been published as *DK-Report No. 2017-04*.

Except for some necessary changes in formatting of the text and equations, positioning of the figures, minor corrections of typing errors and removing of affiliation information, the following chapters are identical to the corresponding published and submitted versions of these articles. All authors contributed equally.

5 Conclusion and outlook

In this thesis we discuss various qualitative aspects of numerical methods for stochastic partial differential equations, where we focus on stability and efficiency.

In Chapter 2 we investigate the interplay of different concepts of stochastic stability and the performance of Monte Carlo estimators. For this we consider linear stochastic differential that are typically arising for spatially discretised SPDEs, of which the corresponding zero solution are asymptotically almost surely stable, but unstable in the mean-square sense. Due to this specific stability setting the Monte Carlo error for estimating the second moment of the solution process can be become arbitrarily large, since the pathwise behaviour of the (numerical) trajectories acts contrarily to the qualitative behaviour of the quantity of interest. Consequently, the Monte Carlo error heavily depends on rarely occurring trajectories that are sufficiently far away from the almost surely attracting equilibrium solution. In order to force these rare events we employ an efficient importance sampling technique that is based on appropriately tuning the upper Lyapunov exponent of the modified linear system of the transformed SDE system.

For future work, an important extension is to investigate the influence of this specific stochastic stability setting for nonlinear stochastic differential equations (with possibly more than a single equilibrium solution). For this we want to emphasize that a structural investigation of the Lyapunov spectrum of linear and nonlinear SODE systems would lead to a better understanding of such effects. As a second extension of Chapter 2 the proposed importance sampling technique could be extended to a more general framework that would allow to use a larger class of measure transformations that also could handle the nonlinear setting e.g. by using transformations that are based on the pathwise behaviour of the solution process.

In Chapter 3, we construct importance sampling techniques for Hilbert space-valued SPDEs by using an infinite-dimensional version of the well-known Girsanov theorem for Q-Wiener processes. In contrast to Chapter 2, where we considered measure transformations of finite-dimensional SODEs (based on semidiscretisated SPDEs) this approach now decouples the importance sampling techniques from spatial discretisation schemes and does not require any specific structure of the semidiscretised system of SDEs. As a guidance for the construction of a large class of measure transformations we show the existence of an optimal measure transformation for semilinear SPDEs. Besides the theoretical justification of the proposed importance sampling techniques we discuss various implementation issues and present numerical experiments showing a remarkable reduction of the Monte Carlo error compared to standard numerical experiments.

A first possible continuation of this work is to extend these results to SPDEs that are driven by more general stochastic processes such as cylindrical Wiener processes or Lévy processes. Due to the fact that the techniques for constructing the infinite-dimensional measure transformations heavily depend on the trace-class property of the covariance operator and on the Gaussian distribution of the increments, this extension requires different or at least more sophisticated methods to reduce the variance of the quantity of interest in numerical experiments.

In Chapter 4, we develop a structural mean-square stability analysis of SPDE approximations. For this we first set up an abstract framework, for which we can show that – as in

the linear SODE setting – the asymptotic mean-square stability properties of the underlying zero solution can be directly related to an eigenvalue problem of a deterministic tensor product space-valued operator. Based on this result we investigate the mean-square stability properties of standard numerical methods for SPDEs such as various Maruyama- and Milstein-type schemes. By deriving sufficient conditions for the asymptotic mean-square stability of the discrete zero solution we illustrate the importance of the used rational approximation for approximating the underlying semigroup. In particular, we are able to prove that under certain conditions the zero solution of the SPDE as well as of numerical schemes based on the backward Euler-Maruyama and the backward Milstein scheme are asymptotically mean-square stable without any restriction on the chosen refinement parameters. Finally, numerical experiments for standard finite element and spectral Galerkin methods illustrate the theoretical results.

A natural extension of this work is to consider the nonlinear setting. We have already presented a promising attempt by adapting Lyapunov function techniques for stochastic recurrences (see Section 2.1 of this introduction), where we considered approximations of semilinear SPDEs with globally Lipschitz continuous operator F and G. However, this would need further investigations to provide a better insight into the nonlinear mean-square stability analysis. Another very interesting extension of the work in Chapter 4 would be to investigate the asymptotic a.s. stability properties of the zero solution of SPDE approximations.

Finally in Chapter 5, we present a combination of space-time multigrid techniques with multilevel Monte Carlo methods. Space-time methods for parabolic PDEs are already well established in the deterministic literature and multilevel Monte Carlo methods are commonly used for SPDEs. However, the combination of both methods is a completely new approach that enables the parallelisation of the problem in space, time and probability. We develop and apply the proposed method in the context of SDEs with additive noise, such as in finite dimensions for the Ornstein-Uhlenbeck process and in infinite dimensions for the stochastic heat equation with additive noise. We conclude by showing the good convergence and parallelisation properties of the proposed algorithm. In particular, we illustrate in numerical experiments for the stochastic heat equation that the best computation times are achieved by using a balanced distribution of cores with respect to the parallelisation in time and the parallelisation of the Monte Carlo estimators.

An open question in this work is how one could extend the analysis of the space-time solver to parabolic SPDEs with linear multiplicative noise. The treatment of multiplicative noise implies that the space-time system matrix (e.g. based on the backward Euler-Maruyama scheme) becomes a matrix with random coefficients in the sub-diagonal. To be more precise the problem for this setting is that we cannot guarantee the convergence and the robustness of the space-time solver by using the same Fourier analysis techniques as presented in [24].

References

[1] A. Abdulle and A. Blumenthal. “Stabilized multilevel Monte Carlo method for stiff stochastic differential equations”. In: *J. Comput. Phys.* 251 (2013), pp. 445–460.

[2] A. Andersson, R. Kruse, and S. Larsson. “Duality in refined Sobolev-Malliavin spaces and weak approximation of SPDE”. In: *Stoch. Partial Differ. Equ. Anal. Comput.* 4.1 (2016), pp. 113–149.

[3] L. Arnold. *Random dynamical systems*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 1998.

[4] L. Arnold. *Stochastic differential equations: theory and applications*. Translated from the German. Wiley-Interscience [John Wiley & Sons], New York-London-Sydney, 1974.

[5] A. Barth and A. Lang. “Milstein approximation for advection-diffusion equations driven by multiplicative noncontinuous martingale noises”. In: *Appl. Math. Optim.* 66.3 (2012), pp. 387–413.

[6] A. Barth and A. Lang. “Multilevel Monte Carlo method with applications to stochastic partial differential equations”. In: *Int. J. Comput. Math.* 89.18 (2012), pp. 2479–2498.

[7] A. Barth and A. Lang. “Simulation of stochastic partial differential equations using finite element methods”. In: *Stochastics* 84.2-3 (2012), pp. 217–231.

[8] A. Barth, A. Lang, and C. Schwab. “Multilevel Monte Carlo method for parabolic stochastic partial differential equations”. In: *BIT* 53.1 (2013), pp. 3–27.

[9] A. Barth, C. Schwab, and N. Zollinger. “Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients”. In: *Numer. Math.* 119.1 (2011), pp. 123–161.

[10] G. Berkolaiko et al. “Almost sure asymptotic stability analysis of the θ-Maruyama method applied to a test system with stabilising and destabilising stochastic perturbations”. In: *LMS J. Comput. Math.* 15 (2012), pp. 71–83.

[11] W.-J. Beyn and R. Kruse. “Two-sided error estimates for the stochastic theta method”. In: *Discrete Contin. Dyn. Syst. Ser. B* 14.2 (2010), pp. 389–407.

[12] T. Bradley et al. “Parallelization techniques for random number generations”. In: *GPU Computing Gems Emerald Edition* 16 (2011), pp. 231–246.

[13] D. Braess. *Finite elements. Theory, fast solvers, and applications in solid mechanics*. Second Edition. Cambridge University Press, Cambridge, 2001.

[14] Z. Brzeźniak et al. “Itô’s formula in UMD Banach spaces and regularity of solutions of the Zakai equation”. In: *J. Differential Equations* 245.1 (2008), pp. 30–58.

[15] E. Buckwar and C. Kelly. “Non-normal drift structures and linear stability analysis of numerical methods for systems of stochastic differential equations”. In: *Comput. Math. Appl.* 64.7 (2012), pp. 2282–2293.

[16] E. Buckwar and C. Kelly. “Towards a systematic linear stability analysis of numerical methods for systems of stochastic differential equations”. In: *SIAM J. Numer. Anal.* 48.1 (2010), pp. 298–321.

[17] E. Buckwar, M. G. Riedler, and P. E. Kloeden. “The numerical stability of stochastic ordinary differential equations with additive noise”. In: *Stoch. Dyn.* 11.2-3 (2011), pp. 265–281.

[18] E. Buckwar and T. Sickenberger. “A structural analysis of asymptotic mean-square stability for multi-dimensional linear stochastic differential systems”. In: *Applied Numerical Mathematics* 62.7 (2012), pp. 842–859.

[19] J. M. C. Clark and R. J. Cameron. “The maximum rate of convergence of discrete approximations for stochastic differential equations”. In: *Stochastic differential systems (Proc. IFIP-WG 7/1 Working Conf., Vilnius, 1978)*. Vol. 25. Lecture Notes in Control and Information Sci. Springer, Berlin-New York, 1980, pp. 162–171.

[20] J. Creutzig et al. “Infinite-dimensional quadrature and approximation of distributions”. In: *Found. Comput. Math.* 9.4 (2009), pp. 391–429.

[21] G. Da Prato, M. Röckner, and A. Jentzen. “A mild Itô formula for SPDEs”. eprint: arXiv:1009.3526 (2011), 31 pages.

[22] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and its Applications, 44. Cambridge University Press, Cambridge, 1992.

[23] J. G. Gaines and T. J. Lyons. “Random generation of stochastic area integrals”. In: *SIAM J. Appl. Math.* 54.4 (1994), pp. 1132–1146.

[24] M. J. Gander and M. Neumüller. “Analysis of a New Space-Time Parallel Multigrid Algorithm for Parabolic Problems”. In: *SIAM J. Sci. Comput.* 38.4 (2016), A2173–A2208.

[25] M. J. Gander and M. Neumüller. “Analysis of a Time Multigrid Algorithm for DG-Discretizations in Time”. eprint: arXiv:1409.5254 (2014), 26 pages.

[26] L. Gawarecki and V. Mandrekar. *Stochastic differential equations in infinite dimensions with applications to stochastic partial differential equations*. Probability and its Applications (New York). Springer, Heidelberg, 2011.

[27] B. Ghayebi and S. M. Hosseini. “A simplified Milstein scheme for SPDEs with multiplicative noise”. In: *Abstr. Appl. Anal.* (2014). Article ID 140849, 15 pages.

[28] M. B. Giles. “Multilevel Monte Carlo methods”. In: *Acta Numer.* 24 (2015), pp. 259–328.

[29] M. B. Giles. “Multilevel Monte Carlo path simulation”. In: *Oper. Res.* 56.3 (2008), pp. 607–617.

[30] M. B. Giles and C. Reisinger. “Stochastic finite differences and multilevel Monte Carlo for a class of SPDEs in finance”. In: *SIAM J. Financial Math.* 3.1 (2012), pp. 572–592.

[31] A. Haghighi, S. M. Hosseini, and A. Rößler. “Diagonally drift-implicit Runge-Kutta methods of strong order one for stiff stochastic differential systems”. In: *J. Comput. Appl. Math.* 293 (2016), pp. 82–93.

[32] A.-L. Haji-Ali et al. “Optimization of mesh hierarchies in multilevel Monte Carlo samplers”. In: *Stoch. Partial Differ. Equ. Anal. Comput.* 4.1 (2016), pp. 76–112.

[33] D. J. Higham, X. Mao, and A. M. Stuart. “Exponential mean-square stability of numerical solutions to stochastic differential equations”. In: *LMS J. Comput. Math.* 6 (2003), pp. 297–313.

[34] P. Imkeller and C. Lederer. “Some formulas for Lyapunov exponents and rotation numbers in two dimensions and the stability of the harmonic oscillator and the inverted pendulum”. In: *Dyn. Syst.* 16.1 (2001), pp. 29–61.

[35] A. Jentzen and M. Röckner. “A Milstein Scheme for SPDEs”. In: *Found. Comput. Math.* 15.2 (2015), pp. 313–362.

[36] A. Kebaier. “Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing”. In: *Ann. Appl. Probab.* 15.4 (2005), pp. 2681–2705.

[37] C. Kelly. “Stochastic stability analysis of a reduced galactic dynamo model with perturbed α-effect”. In: *Phys. A* 457 (2016), pp. 480–491.

[38] R. Khasminskii. *Stochastic Stability of Differential Equations*. Stochastic Modelling and Applied Probability, 66. Springer, Heidelberg, 2012.

[39] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Applications of Mathematics, 23. Springer-Verlag, Berlin, 1992.

[40] P. E. Kloeden, E. Platen, and I. W. Wright. “The approximation of multiple stochastic integrals”. In: *Stochastic Anal. Appl.* 10.4 (1992), pp. 431–441.

[41] V. Kolmanovskii and L. Shaikhet. “General method of Lyapunov functionals construction for stability investigation of stochastic difference equations”. In: *Dynamical systems and applications*. Vol. 4. World Sci. Ser. Appl. Anal. World Sci. Publ., River Edge, NJ, 1995, pp. 397–439.

[42] R. Kruse. “Consistency and stability of a Milstein-Galerkin finite element scheme for semilinear SPDE”. In: *Stoch. Partial Differ. Equ. Anal. Comput.* 2.4 (2014), pp. 471–516.

[43] R. Kruse. *Strong and Weak Approximation of Semilinear Stochastic Evolution Equations*. Lecture Notes in Mathematics, 2093. Springer, Cham, 2014.

[44] A. Lang. “A note on the importance of weak convergence rates for SPDE approximations in multilevel Monte Carlo schemes”. In: *Monte Carlo and quasi-Monte Carlo methods*. Vol. 163. Springer Proc. Math. Stat. Springer, [Cham], 2016, pp. 489–505.

[45] A. Lang and A. Petersson. “Monte Carlo versus multilevel Monte Carlo in weak error simulations of SPDE approximations”. In: *Mathematics and Computers in Simulation* (2017).

[46] A. Lang and J. Potthoff. “Fast simulation of Gaussian random fields”. In: *Monte Carlo Methods Appl.* 17.3 (2011), pp. 195–214.

[47] K. Liu. *Stability of Infinite Dimensional Stochastic Differential Equations with Applications*. Chapman & Hall/CRC Monographs, Surveys in Pure, and Applied Mathematics 135, 2006.

[48] W. Liu and M. Röckner. *Stochastic partial differential equations: an introduction*. Universitext. Springer, Cham, 2015.

[49] G. J. Lord, C. E. Powell, and T. Shardlow. *An Introduction to Computational Stochastic PDEs*. Cambridge Texts in Applied Mathematics. Cambridge University Press, New York, 2014.

[50] V. Mandrekar and B. Rüdiger. *Stochastic integration in Banach spaces*. Vol. 73. Probability Theory and Stochastic Modelling. Theory and applications. Springer, Cham, 2015.

[51] X. Mao. *Stochastic Differential Equations and their Applications*. Horwood Publishing Series in Mathematics & Applications. Horwood Publishing Limited, Chichester, 1997.

[52] G. N. Milstein, E. Platen, and H. Schurz. “Balanced implicit methods for stiff stochastic systems”. In: *SIAM J. Numer. Anal.* 35.3 (1998), pp. 1010–1019.

[53] G. N. Milstein and M. V. Tretyakov. “Practical variance reduction via regression for simulating diffusions”. In: *SIAM Journal on Numerical Analysis* 47.2 (2009), pp. 887–910.

[54] G. N. Milstein and M. V. Tretyakov. “Solving parabolic stochastic partial differential equations via averaging over characteristics”. In: *Math. Comp.* 78.268 (2009), pp. 2075–2106.

[55] G. N. Milstein and M. V. Tretyakov. *Stochastic Numerics for Mathematical Physics*. Scientific Computation. Springer-Verlag, Berlin, 2004.

[56] B. Øksendal. *Stochastic differential equations. An introduction with applications*. Fourth Edition. Universitext. Springer-Verlag, Berlin, 1995.

[57] B. Paternoster and L. Shaikhet. “About stability of nonlinear stochastic difference equations”. In: *Appl. Math. Lett.* 13.5 (2000), pp. 27–32.

[58] S. Peszat and J. Zabczyk. *Stochastic Partial Differential Equations with Lévy Noise. An Evolution Equation Approach*. Vol. 113. Encyclopedia of Mathematics and Its Applications. Cambridge University Press, 2007.

[59] C. Prévôt and M. Röckner. *A concise course on stochastic partial differential equations*. Vol. 1905. Lecture Notes in Mathematics. Springer, Berlin, 2007.

[60] P. E. Protter. *Stochastic integration and differential equations*. Second Edition. Vol. 21. Stochastic Modelling and Applied Probability. Springer-Verlag, Berlin, 2005.

[61] C. Reisinger. “Mean-square stability and error analysis of implicit time-stepping schemes for linear parabolic SPDEs with multiplicative Wiener noise in the first derivative”. In: *Int. J. Comput. Math.* 89.18 (2012), pp. 2562–2575.

[62] L. E. Shaikhet. “Construction of Lyapunov functionals for stochastic difference equations with continuous time”. In: *Math. Comput. Simulation* 66.6 (2004), pp. 509–521.

[63] T. Shardlow. “Numerical methods for stochastic parabolic PDEs”. In: *Numer. Funct. Anal. Optim.* 20.1-2 (1999), pp. 121–145.

[64] A. M. Stuart and A. R. Humphries. *Dynamical systems and numerical analysis*. Vol. 2. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge, 1996.

[65] D. Talay. “Approximation of upper Lyapunov exponents of bilinear stochastic differential systems”. In: *SIAM J. Numer. Anal.* 28.4 (1991), pp. 1141–1164.

[66] D. Talay and L. Tubaro. “Expansion of the global error for numerical schemes solving stochastic differential equations”. In: *Stochastic Anal. Appl.* 8.4 (1990), pp. 483–509.

[67] S. Tappe. “Some refinements of existence results for SPDEs driven by Wiener processes and Poisson random measures”. In: *Int. J. Stoch. Anal.* (2012), Art. ID 236327, 24 pages.

[68] V. Thomée. *Galerkin Finite Element Methods for Parabolic Problems*. 2nd. Vol. 25. Springer Series in Computational Mathematics. Springer, 2006.

[69] M. Wiktorsson. “Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions”. In: *Ann. Appl. Probab.* 11.2 (2001), pp. 470–487.

2 An importance sampling technique in Monte Carlo methods for SDEs with a.s. stable and mean-square unstable equilibrium

M. Ableidinger, E. Buckwar and A. Thalhammer

In this work we investigate the interplay of almost sure and mean-square stability for linear SDEs and the Monte Carlo method for estimating the second moment of the solution process. In the situation where the zero solution of the SDE is asymptotically stable in the almost sure sense but asymptotically mean-square unstable, the latter property is determined by rarely occurring trajectories that are sufficiently far away from the origin. The standard Monte Carlo approach for estimating higher moments essentially computes a finite number of trajectories and is bound to miss those rare events. It thus fails to reproduce the correct mean-square dynamics (under reasonable cost). A straightforward application of variance reduction techniques will typically not resolve the situation unless these methods force the rare, exploding trajectories to happen more frequently. Here we propose an appropriately tuned importance sampling technique based on Girsanov's theorem to deal with the rare event simulation. In addition further variance reduction techniques, such as multilevel Monte Carlo, can be applied to control the variance of the modified Monte Carlo estimators. As an illustrative example we discuss the numerical treatment of the stochastic heat equation with multiplicative noise and present simulation results.

Key words. Monte Carlo estimators, Systems of linear SDEs, Asymptotic almost sure stability, Asymptotic mean-square stability, Variance reduction, Rare event simulation

AMS subject classifications. 60H10, 65C05, 65M12, 60H15

1 Introduction

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a complete probability space, where the filtration $(\mathcal{F}_t)_{t\in[0,T]}$ satisfies the usual conditions. We consider d-dimensional systems of linear Itô stochastic differential equations

$$\mathrm{d}X(t) = AX(t)\,\mathrm{d}t + \sum_{r=1}^{m} G_r X(t)\,\mathrm{d}\beta_r(t), \qquad X(0) = X_0, \quad t \in [0,T], \tag{1}$$

where $A, G_r \in \mathbb{R}^{d\times d}$, β_r are m independent, real-valued Wiener processes, and the initial value X_0 is in $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. Under these assumptions Eq. (1) has a unique strong solution, which, however, allows an explicit representation only for special cases such as for systems with commuting matrices A, G_r, see [2, Section 8.5, Remark 8.5.9]. For the case of a zero initial value $X_0 = 0$, Eq. (1) admits the zero solution $X_e(t) \equiv 0$ for all t.

In this article we are interested in the numerical approximation of quantities of the type $\mathbb{E}f(X(t))$ for some functional $f : \mathbb{R}^d \to \mathbb{R}$, which has at least quadratic growth in $|x|$.

We first describe the setting of Monte Carlo techniques and error analysis in terms of interpreting

$$E_M[f(\widehat{X}(t)] := \frac{1}{M}\sum_{i=1}^{M} f(\widehat{X}^{(i)}(t)) \tag{2}$$

as an estimator for $\mathbb{E}f(X(t))$. Here, $\widehat{X}^{(i)}(t)$ denotes an independent realisation of the approximated solution $\widehat{X}(t)$, which we obtain by using a numerical integrator of weak order p. The approximation of the trajectories produces a *systematic* error, which can be expressed as ([19, 22])

$$\mathbb{E}[f(X(t))] = \mathbb{E}[f(\widehat{X}(t))] + \mathcal{O}(h^p).$$

Subsequently we will assume that the time step size is sufficiently small such that the systematic error is dominated by the *Monte Carlo* error, which can be expressed as ([19])

$$E_M[f(\widehat{X}(t))] = \frac{1}{M}\sum_{i=1}^{M} \mathbb{E}f(\widehat{X}^{(i)}(t)) \pm c\frac{\mathsf{Var}[f(\widehat{X}(t)))]^{1/2}}{M^{1/2}}.$$

Under the assumption that the variance of the approximated random variable $f(\widehat{X}(t))$ is close to the variance of $f(X(t))$, the quality of the Monte Carlo estimator depends on the variance of the underlying SDE. Thus for problems with large variances one needs either a large number of realisations M, i.e. a very fine discretisation of the underlying probability space, or estimators with a smaller variance than standard Monte Carlo estimators. This second consideration directly leads to the field of variance reduction techniques, for an overview see for example [8, 3].

In this work we analyse the impact of long time properties of linear systems of SDEs on Monte Carlo estimators, in particular how different concepts of stochastic stability, i.e. asymptotic stability in the almost sure and mean-square sense (see Definition 2.1), affect the numerical results using the standard Monte Carlo estimator defined in Eq. (2).

Ignoring any systematic error for the moment, we choose the following trivial observation as a starting point for our considerations: The estimator defined in Eq. (2) consists only of finitely many realisations of the solution process, and therefore the long time evolution is heavily influenced by the path-wise stability properties of the SDE. However, due to the (at least) quadratic growth of $f(x)$, $\mathbb{E}f(X(t))$ is governed by the corresponding mean-square stability of the zero solution of (1). Hence, the characterisation of our problem can be

summarised by using the following observations:

I The zero solution of system (1) can be asymptotically stable in the almost sure sense but at the same time asymptotically mean-square unstable. This situation is well known in the literature, e.g. for the geometric Brownian motion (see [16] and Section 2).

II An immediate consequence of the above situation is that the exponential growth in time of $\mathbb{E}|X(t)|^2$ is due to very rare exploding trajectories. This leads to a prohibitively high number of realisations needed for the standard Monte Carlo estimation to obtain a decent approximation of the second or higher moments of the solution process $X(t)$.

III The problem is not purely academic: Space discretising diffusion-type SPDEs with e.g. finite differences leads to high dimensional SODE systems of type (1). The solution trajectories decrease rapidly due to the dissipative properties of the Laplacian and the stabilising structure of the diffusion matrices G_r. At the same time, the equilibrium solution will become asymptotically mean-square unstable for some, often moderate, value of noise intensity and increasing the noise intensity will amplify this effect. Note that the computational cost for simulating such high-dimensional SDE systems automatically prohibits a substantial increase of the number of trajectories.

IV A straightforward application of multilevel Monte Carlo techniques for reducing the variance will typically not resolve the situation, unless these methods force the rare exploding trajectories to happen more frequently.

We propose the following modification of Monte Carlo estimation by importance sampling techniques: We change the drift of the system in such a way that the trajectories explode more often. This can be achieved by transforming the underlying probability measure due to Girsanov's theorem. Weighting the trajectories with the corresponding density process results in an unbiased Monte Carlo estimator for the desired quantity. An important property of the developed measure transformation is that it can be precomputed and does not change the linearity of the system. Consequently, the computational cost of our proposed method is the same as for the standard Monte Carlo estimation. Further, the variance of the modified estimators can also be reduced by applying additional variance reduction techniques such as multilevel Monte Carlo methods.

In Section 2 we will give a short review of necessary notions of stability theory of stochastic differential equations and we discuss the points I and II by providing analytical and numerical results for the geometric Brownian motion. In Section 2.1 we introduce a spatially discretised stochastic heat equation as an illustrative and non-trivial example and discuss the stability properties of the zero solution of the resulting system of SODEs. In Section 3 we discuss different variance reduction techniques (multilevel Monte Carlo methods and importance sampling) for the spatially discretised stochastic heat equation and

formulate our proposed method. We will present numerical results in Section 4 and close this work with conclusions in Section 5.

2 Stability theory

In this section we summarise the main definitions and notation from stochastic stability theory and illustrate the differences between these stability concepts by considering geometric Brownian motion.

We treat the following two types of (asymptotic) stochastic stability of the equilibrium solution of Eq. (1), which is given by the zero solution $X_e(t) \equiv 0$ for all t, see [16, 13].

Definition 2.1. (i) The zero solution is called *asymptotically almost sure stable*

(a) if for any given $0 < \varepsilon < 1, \varepsilon' > 0$, there exists $\delta > 0$ such that if $|X_0| < \delta$, then $\mathbb{P}(|X(t)| > \varepsilon') < \varepsilon$ for all $t \geq 0$ and

(b) if there exists a $\delta' > 0$ such that for any X_0, satisfying $|X_0| < \delta'$ a.s., it holds that $\lim_{t\to\infty} |X(t)| = 0$ a.s.

(ii) The zero solution is called *asymptotically mean-square stable*

(a) if for $\mathbb{E}|X_0|^2 \leq \delta$, then $\sup_{t\geq s} \mathbb{E}|X(t)|^2 \to 0$ as $\delta \to 0$ and

(b) if there exists a $\delta' > 0$ such that for any X_0, satisfying $\mathbb{E}|X_0|^2 < \delta'$, it holds that $\lim_{t\to\infty} \mathbb{E}|X(t)|^2 = 0$.

In order to illustrate the difference between asymptotically almost sure stable and asymptotically mean-square stable equilibria, we recapitulate the well-known results for the geometric Brownian motion (see [2, 15]). Let

$$\mathrm{d}X(t) = \lambda X(t)\,\mathrm{d}t + \sigma X(t)\,\mathrm{d}\beta(t), \quad X(0) = X_0, \tag{3}$$

where $\lambda, \sigma, X_0 \in \mathbb{R}$. The explicit solution and the second moment of Eq. (3) are given by

$$X(t) = X_0 \exp\left(\left(\lambda - \frac{\sigma^2}{2}\right) t + \sigma\beta(t)\right), \tag{4}$$

$$\mathbb{E}X^2(t) = X_0^2 \exp\left(\left(2\lambda + \sigma^2\right) t\right). \tag{5}$$

Obviously the zero solution is asymptotically mean-square stable if and only if $2\lambda + \sigma^2 < 0$. Using the law of iterated logarithms, it can be seen that the zero solution is asymptotically stable in the almost sure sense if and only if $\lambda - \sigma^2/2 < 0$. As a consequence, the zero solution of Eq. (3) becomes unavoidably asymptotically stable in the almost sure and asymptotically unstable in the mean-square sense for sufficiently large σ (see also the discussion in [16, Example 1.4.1]).

Suppose we want to approximate $\mathbb{E}X^2(t)$ by standard Monte Carlo estimation, where we use the exact solution (4) for simulating the paths, i.e. no systematic error arises. For a

given set of parameters we calculate the probability that $X^2(t)$ stays in an ϵ-neighbourhood of 0 with a certain probability $1-\alpha$. We get

$$\mathbb{P}\left[X^2(t) < \epsilon\right] = 1-\alpha \iff \mathbb{P}\left[\xi < \frac{\ln(\epsilon/X_0^2) - 2(\lambda - \sigma^2/2)t}{2\sigma\sqrt{t}}\right] = 1-\alpha$$

for $\xi \sim \mathcal{N}(0,1)$. Note that for a fixed probability $1-\alpha$ the bound ϵ *decreases* exponentially for growing σ and t, whereas the mean-square process (5) *grows* exponentially in σ and t. This compromises the results of Monte Carlo estimation and we illustrate this with the following parameter setting: Let $\lambda = -2$, $\sigma = 3$, and $X_0 = 1$, so the zero solution is asymptotically mean-square unstable. We further fix the time at $T = 10$ and obtain that the analytic solution of the mean-square process (5) has order of magnitude

$$\mathbb{E}X^2(T) \approx 10^{21}.$$

However, it is highly improbable to sample values substantially larger than zero. The probability that all paths of a standard Monte Carlo estimation with 10^4 trajectories are almost zero within machine accuracy is $1-10^{-3}$, since

$$\mathbb{P}\left[\left(X^{(i)}(T)\right)^2 < 10^{-14}\right] \approx 1 - 10^{-7}$$

and

$$\prod_{i=1}^{10^4} \mathbb{P}\left[\left(X^{(i)}(T)\right)^2 < 10^{-14}\right] \approx 1 - 10^{-3}.$$

As a consequence, the standard Monte Carlo estimator $E_M[X^2(T)]$ fails to approximate $\mathbb{E}X^2(T)$ (see Figure 1).

This effect is also present in the case of higher dimensional SODE systems (see Figure 2). In the next section, we consider this situation by using a space-discretised version of a stochastic heat equation with multiplicative noise.

2.1 Linear systems of SODEs

In this section we present the semi-discretised stochastic heat equation as a benchmark problem for spatially discretised diffusion type SPDEs and analyse the qualitative behaviour of the zero solution.

Stochastic heat equation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space with filtration $(\mathcal{F}_t)_{t\in[0,T]}$. We consider the one-dimensional stochastic heat equation on the spatial domain $\mathcal{D} = [0,1]$ with homogenous

Dirichlet conditions:

$$\begin{aligned} \mathrm{d}u(x,t) &= \Delta u(x,t)\,\mathrm{d}t + \sigma u(x,t)\,\mathrm{d}W(x,t), \\ u(0,\cdot) &= u(1,\cdot) = 0, \\ u(x,0) &= u_0(x) = \sin(\pi x) \quad \text{for } x \in \mathcal{D}, \end{aligned} \tag{6}$$

with a sufficiently smooth initial function, since $u_0 \in L^2(\mathcal{D})$. The multiplicative noise is interpreted in the sense of Nemytskii operators (see [17]). For the driving noise we assume that $W(x,t)$ is an $\mathcal{F}_t$-adapted Q-Wiener process with a linear, non-negative and symmetric covariance operator Q. Furthermore Q is of trace class and has an orthonormal basis of eigenfunctions $\{\chi_j, j \in \mathbb{N}\}$ with eigenvalues $q_j \geq 0$. Under these assumptions $W(x,t)$ can be represented in the following series expansion

$$W(x,t) = \sum_{j=1}^{\infty} \sqrt{q_j}\chi_j(x)\beta_j(t), \tag{7}$$

which converges in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ (see [5]). Here, $\beta_j(t)$ are independent one-dimensional Wiener processes. The existence of a unique mild solution of (6) is guaranteed by the classical framework of [5].

Space discretisation and model problem

For spatial discretisation we apply the standard finite difference scheme on an equidistant spatial mesh $\{x_0, \ldots, x_{N+1}\}$ with mesh width $h = 1/(N+1)$. Then the spatially discretised version of Eq. (6) takes the form

$$\mathrm{d}u(x_i,t) = \left(\Delta_h u(t)\right)(x_i)\,\mathrm{d}t + \sigma u(x_i,t)\,\mathrm{d}W(x_i,t),$$

where Δ_h denotes the three-point discrete Laplacian. Using the eigenfunctions of the Laplacian as a basis representation for $W(x,t)$ we define the following truncation of the series (7) (see [17]),

$$W^N(x,t) := \sum_{i=1}^{N} \sqrt{q_i}\chi_i(x)\beta_i(t)$$

with $q_i = 2i^{-(2r+1+\epsilon)}$ and $\chi_i(x) = \sqrt{2}\sin(\pi i x)$ for all $x \in \mathcal{D}$. Here $\epsilon > 0$ and r controls the regularity of the Wiener process in the sense that $W(t) \in H_0^r(0,1)$. Denoting the solution-vector at the interior points of the spatial mesh as

$$X(t) = (X_1(t), \ldots X_N(t))^T = (u(x_1,t), \ldots, u(x_N,t))^T,$$

we obtain the following N-dimensional SODE-system

$$\begin{aligned} \mathrm{d}X(t) &= A^h X(t)\,\mathrm{d}t + \sum_{i=1}^{N} G_i^h X(t)\,\mathrm{d}\beta_i(t), \\ X(0) &= (\sin(x_1\pi), \ldots, \sin(x_N\pi))^T, \end{aligned} \tag{8}$$

where

$$A^h = \frac{1}{h^2}\,\mathrm{tridiag}\left(1, -2, 1\right)_{j=1}^{N}$$

and

$$G_i^h = \mathrm{diag}\left\{\sqrt{2q_i}\sin\left(\frac{j\pi i}{N+1}\right)\right\}_{j=1}^{N}.$$

2.2 Linear stability analysis of Eq. (8)

The mean-square stability of the zero solution for linear SODE systems can be characterised via the *mean-square stability matrix* of the system. Following [4], the process $Y(t) = \mathrm{vec}(X(t)X(t)^T)$ satisfies the deterministic ODE

$$\mathrm{d}\mathbb{E}(Y(t)) = S\mathbb{E}(Y(t))\,\mathrm{d}t, \tag{9}$$

where

$$\begin{aligned} S &= A^h \otimes \mathbb{I}_N + \mathbb{I}_N \otimes A^h + \sigma^2 \sum_{r=1}^{N} G_r^h \otimes G_r^h \\ &= A^h \oplus A^h + \sigma^2 \sum_{r=1}^{N} G_r^h \otimes G_r^h. \end{aligned}$$

Here, $\otimes$ denotes the matrix Kronecker product, $\oplus$ denotes the Kronecker sum and $\mathbb{I}_N$ denotes the N-dimensional identity matrix.

Based on the following lemma, see e.g. [4], we can determine the mean-square stability properties of the zero solution of Eq. (8) by considering the spectrum of the underlying mean-square stability matrix S.

Lemma 2.2. *Let $\alpha(S)$ be the spectral abscissa of S. The zero solution of Eq.* (8) *is asymptotically mean-square stable if and only if $\alpha(S) < 0$.*

Both $A^h \oplus A^h$ and $\sum G_i^h \otimes G_i^h$ are symmetric, so we can apply Weyl's theorem, see e.g. [9, Theorem 4.3.1], to obtain a lower bound for $\alpha(S)$:

$$\max_{1\le j\le N^2}\left[\lambda_{N^2-j+1}\left(A^h \oplus A^h\right) + \sigma^2\lambda_j\left(\sum_{i=1}^{N} G_i^h \otimes G_i^h\right)\right] \le \alpha(S).$$

Here, $\lambda_j(M)$ denotes the j-th largest eigenvalue of a matrix M. Due to basic properties of the Kronecker product (see e.g. [21]) and the form of G_i^h, the matrix $\sum G_i^h \otimes G_i^h$ has

positive eigenvalues. Consequently, the zero solution of Eq. (6) is asymptotically mean-square unstable for large enough σ.

In analogy to the one dimensional case we expect that there exist values of σ for which the zero solution is asymptotic a.s. stable but asymptotic mean-square unstable. In fact, the authors in [15] showed that for linear SODE systems the asymptotic almost sure stability can be interpreted as the limiting case of p-th moment stability for $p \to 0$. However the computation of this limit is not straightforward and the p-th moments for $p < 1$ are hard to interpret. An alternative way for showing asymptotic almost sure stability of the zero solution is the analysis of the corresponding Lyapunov exponents. The *top Lyapunov exponent* is defined as

$$\Lambda = \lim_{t\to\infty} \frac{1}{t} \log |X(t)| \,.$$

As in the deterministic case, the path-wise stability of an SODE system is completely described by Λ: the zero solution of Eq. (6) is asymptotically almost sure stable iff $\Lambda < 0$ (see [2]). The actual computation of Λ requires a thorough understanding of the exact solution $X(t)$, which is typically not available, consequently analytic results are only known for low dimensional systems (see [10, 12]). For higher dimensions the analysis of the top Lyapunov exponent is still tractable if one assumes non-degeneracy of the driving noise (see [13, 1, 18] and for numerical approximations of Λ based on path simulation [22]). Let $D(\lambda) = \sum_{i=1}^{N} G_i^h \lambda\lambda^T G_i^h \in \mathbb{R}^{N\times N}$ for $\lambda \in \mathbb{R}^N$. We say the noise is non-degenerate if there is a $C > 0$ such that

$$\langle D(\lambda)\zeta, \zeta\rangle \geq C|\lambda|^2|\zeta|^2 \qquad \forall \lambda, \zeta \in \mathbb{R}^N \tag{10}$$

or equivalently, that the dimension of the linear hull of $\{G_1^h x, \ldots, G_N^h x\}$ is equal to N (see [13, Appendix A]). Under Condition (10) the top Lyapunov exponent can be represented as (see [13, Theorem 6.11])

$$\Lambda = \int_{\mathcal{S}^{N-1}} \mathcal{K}(\lambda)\nu(d\lambda) \tag{11}$$

with

$$\mathcal{K}(\lambda) = \left\langle A^h\lambda, \lambda\right\rangle + \frac{1}{2}\operatorname{trace}\left[D(\lambda)\right] - \langle D(\lambda)\lambda, \lambda\rangle \,.$$

Here $\mathcal{S}^{N-1}$ denotes the $(N-1)$-dimensional unit sphere and ν denotes the invariant measure of the process $X(t)/|X(t)|$. The advantage of representation (11) is that even if the integral can not be calculated explicitly, one can analyse the kernel $\mathcal{K}(\lambda)$ to obtain sufficient conditions for asymptotic a.s. (in-)stability of the zero solution.

To use this representation we have to ensure that condition (10) is fulfilled. The diffusion matrices G_k^h are defined via the basis representation of the Laplacian. Consequently the noise is non-degenerate if the number of independent Wiener processes is at least equal

to the dimension of the system, which is the case for Eq. (8).

As the diffusion matrices G_i^h are diagonal and since

$$\operatorname{trace}\left[D(\lambda)\right] = \sigma^2 \left\langle \sum_{k=1}^{N} \left(G_k^h\right)^2 \lambda, \lambda \right\rangle,$$

we can write the kernel $\mathcal{K}(\lambda)$ as $\mathcal{K}(\lambda) = \langle K(\lambda)\lambda, \lambda\rangle$ with

$$K(\lambda) = A^h + \frac{\sigma^2}{2} \sum_{k=1}^{N} (G_k^h)^2 - D(\lambda). \tag{12}$$

By careful inspection of the eigenvalues of $K(\lambda)$, we expect $\mathcal{K}(\lambda)$ to be negative on $\mathcal{S}^{N-1}$ independently of σ. Therefore the zero solution of Eq. (8) is asymptotically a.s. stable, whereas it becomes asymptotically mean-square unstable for σ large enough. Then again the *explosion* of $\mathbb{E}|X(t)|^2$ depends on very rare trajectories and standard Monte Carlo estimation has the same difficulties to approximate $\mathbb{E}|X(t)|^2$ as in the one dimensional case, see Figure 2. In the next section we will treat the question if the mean-square stability behaviour can be recovered in Monte Carlo simulations by using variance reduction techniques.

3 Variance reduction techniques and rare event simulations

As we have already mentioned in Section 1, the Monte Carlo error can be controlled by either increasing the number of samples or by using variance reduction techniques to obtain estimators with smaller variances compared to standard Monte Carlo methods. The first approach, i.e. increasing the number of trajectories, faces the following severe limitations:

1 A typical indicator that the number of simulated random variables might be insufficient is a substantial empirical variance of the standard Monte Carlo estimator. However, in our setting the rapid decay of the paths leads to a nearly vanishing *empirical* variance. Hence any estimation for the necessary number of simulated trajectories has to rely on the a priori knowledge of the rareness of the exploding trajectories. In general, the probability of these rare events is not known. Moreover, in situations where one can calculate the probability of these rare events, see e.g. geometric Brownian motion in Section 2, the probability of the occurrence of these trajectories is so low that the number of required trajectories is unreasonably high.

2 A natural bound on the number of trajectories is imposed by the computational cost of the time integration method. This limits the possibility of increasing the number of numerical trajectories for high dimensional SODE systems.

Concerning variance reduction techniques we distinguish two types of methods. First, there are methods, such as Control Variates and multilevel Monte Carlo approaches, that reduce the variance by adding suitable control quantities to the standard estimators. The

probability space and the distribution of the simulated random variables are unchanged. Thus problem II of the standard Monte Carlo estimator, i.e. the absence of rare events, is still present for this type of techniques. To see this, we examine the multilevel Monte Carlo estimator in Section 3.1 and analyse its disadvantages for our setting.

Second, there are methods, such as importance sampling, that change the underlying probability measure and therefore the distribution of the simulated trajectories. By being appropriately tuned, these methods are able to enforce the rare event of exploding trajectories. In Section 3.2, we present an importance sampling technique based on Girsanov's theorem where we modify the drift such that we simulate realisations of SODE systems with slower decaying trajectories. With the choice of constant weight functions in the Girsanov transformation, the resulting system is still linear and the computational cost to obtain a single realisation is of the same order as for the original system. Note that we can also apply additional variance reduction techniques of the first type to the modified estimator. In the numerical illustrations at the end of Section 4, we choose a combination of the proposed importance sampling technique (for rare event simulation purposes) and a multilevel Monte Carlo approach (for additional variance reduction).

3.1 Multilevel Monte Carlo methods and their failure to solve Problem II

In this section, we consider the multilevel Monte Carlo (MLMC) estimator introduced by [11, 7] as a variance reduction technique. For this, we denote by $f(\widehat{X}_\ell(t)), \ell = 0, \dots, L$, an approximation of $f(X(t))$ using a numerical time integration method on a geometrical hierarchy of time grids with time step sizes $\Delta_\ell = 2^{-l-\kappa}$. Here, we use the parameter κ to ensure that the numerical method is stable on the coarsest level. Furthermore, let $f(\widehat{X}_{-1}(t)) = 0$.

Then, the multilevel Monte Carlo estimator is defined (see e.g. [7, 6])

$$E^L[f(\widehat{X}_L(t))] := \sum_{\ell=0}^{L} E_{N_\ell}[f(\widehat{X}_\ell(t)) - f(\widehat{X}_{\ell-1}(t))],$$

where N_ℓ independent realisations of $f(\widehat{X}_\ell(t)) - f(\widehat{X}_{\ell-1}(t))$ are used for the standard Monte Carlo estimator on each level.

There is a large literature on problems to which the MLMC approach has successfully been applied, see e.g. [6] for a survey. In many of these applications, a remarkable reduction of computational complexity compared to the standard Monte Carlo estimation is observed and a variance reduction of the underlying estimator is achieved.

In [7], a result on the optimal choice of levels L and of numbers of realisations $N_\ell, \ell = 0, \dots, L$, used on each of these levels is given for the multilevel Monte Carlo estimator. However, the number of realisations depends on the variance of the standard Monte Carlo estimators on each level, i.e. N_ℓ depends on the constant $C_\ell > 0$ in the estimate

$$\mathrm{Var}\Big[E_{N_\ell}[f(\widehat{X}_\ell(t)) - f(\widehat{X}_{\ell-1}(t))]\Big] \leq C_\ell 2^{-\ell} \tag{13}$$

for all $\ell = 0, \dots, L$. By considering the variance of the Monte Carlo estimator on level 0, we obtain

$$\mathrm{Var}(E_{N_0}[f(\widehat{X}_0(t))]) = \frac{1}{N_0^2} \sum_{i=1}^{N_0} \mathrm{Var}(f(\widehat{X}_0^{(i)}(t))) = \frac{1}{N_0} \mathrm{Var}(f(\widehat{X}_0^{(1)}(t))).$$

Given an appropriately chosen initial time step size Δt_0, $\mathrm{Var}(f(\widehat{X}_0(t)))$ can be assumed to be close to $\mathrm{Var}(f(X_0(t)))$. Consequently, for SODEs where the zero solution is stable in the almost sure sense but mean-square unstable, we can deduce that the constant C_0 becomes unreasonably large by the same considerations as for the standard Monte Carlo estimator. Thus, we also have for the multilevel Monte Carlo estimator the performance problem that we need prohibitively many realisations (at least on the coarsest level) to obtain a certain accuracy. Although a reduction of computational cost is achieved by using the coarser time grids compared to the standard Monte Carlo approach, it is still not possible to simulate so many realisations that the rare trajectories for reproducing the mean-square instability occur with sufficient frequency.

In the standard MLMC algorithm proposed in [7, 6], the number of realisations on each level is computed by using the optimal choice of N_ℓ based on empirical estimators for the variance of $f(\widehat{X}_\ell(t)) - f(\widehat{X}_{\ell-1}(t))$. However, the empirical estimators for the variance are essentially zero due to the gap between almost sure and mean-square dynamics. Thus a straightforward application of multilevel Monte Carlo techniques without focusing on rare event simulation is not an appropriate approach in our setting. For this reason, we propose below a rare event simulation procedure based on importance sampling where we force the exploding trajectories to happen more frequently. Afterwards we can improve the modified estimator by further variance reduction techniques, e.g. by multilevel Monte Carlo methods as above.

3.2 Importance sampling

The main idea of importance sampling (see e.g. [8, 3, 14]) is to change the underlying probability measure (and therefore the distribution of $X(t)$) in order to either reduce the variance of the random variable $f(X(t))$ or in rare event simulations, to increase the frequency of rare events in the Monte Carlo simulation.

Here we concentrate on the second aspect and transform the underlying probability measure in a delicate way such that the exploding trajectories happen more often, but the zero solution is still asymptotically a.s. stable.

A convenient way to transform the underlying probability measure is given by Girsanov's theorem (see [20]) which allows us to change the drift of the system. Let

$$\mathrm{d}\widetilde{X}(t) = \widetilde{A}^h \widetilde{X}(t)\,\mathrm{d}t + \sigma \sum_{i=1}^{N} G_i^h \widetilde{X}(t)\,\mathrm{d}\beta_i(t) \tag{14}$$

be a (possibly non-linear) system of SODEs with modified drift

$$\widetilde{A}^h = A^h - \sigma \sum_{i=1}^{N} d_i(t, \widetilde{X}(t)) G_i^h$$

for a set of real-valued functions $d_j : [0,T] \times \mathbb{R}^d \to \mathbb{R}$. We define a new probability measure $\widetilde{\mathbb{P}}$ via the density process

$$\mathrm{d}\Theta(t) = \sum_{j=1}^{N} d_j(t, \widetilde{X}(t))\, \Theta(t)\, \mathrm{d}\beta_j(t). \tag{15}$$

Under certain conditions on the functions d_j, e.g. they fulfill Novikov's condition (see [20]) we can apply Girsanov's theorem, see e.g. [20, Theorem 46], which states that $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equivalent and the processes

$$\gamma_i(t) = -\int_0^t d_i(s, \widetilde{X}(s))\, \mathrm{d}s + \beta_i(t)$$

are Wiener processes under $\widetilde{\mathbb{P}}$. Consequently we have

$$\mathbb{E}\left[f(X(t))\right] = \mathbb{E}\left[f(\widetilde{X}(t))\Theta(t)\right]$$

and we can use trajectories of Eq. (14) to estimate the quantity $\mathbb{E}\left[f(X(t))\right]$.
Concerning the choice of the weight functions d_j we stipulate the following conditions:

- The rare events of exploding trajectories should happen more frequently.
- The computational effort for simulating Eq. (14) should not be larger than that for simulating Eq. (8).

It is well-known in the literature (see [19]) that there exists an optimal choice $d_j^*(t, \widetilde{X}(t))$ for the weight functions for which the quantity $f(\widetilde{X}(t))\Theta(t)$ becomes deterministic and consequently the Monte Carlo estimator has variance 0.

Of course the optimal value is typically not known, and approximated solutions of the corresponding Kolmogorov backward equation (as proposed in [19]) are non-trivial to obtain. Additionally any choice of $d_j(t, \widetilde{X}(t))$ which is not constant results in a possibly high-dimensional system of non-linear SODEs leading to additional numerical challenges as e.g. to ensure the non-negativity of the density process $\Theta(t)$. To circumvent this problem we propose using only constant weight functions d_j, which implies that the system (14) is linear and the density process $\Theta(t)$ can be calculated explicitly.

Remark 3.1. In special cases, the optimal choice of weight functions $d_j^*(t, \widetilde{X}(t))$ for the second moment are constants, e.g. for the geometric Brownian motion (3), where

$$\widetilde{X}^2(t)\Theta(t) = X_0^2 \exp\left[2\left(\lambda - d\sigma - \frac{\sigma^2}{2}\right)t - \frac{d^2}{2}t + (2\sigma + d)\,\beta(t)\right] \tag{16}$$

is deterministic for $d^* = -2\sigma$.

For multi-dimensional linear systems of SODEs with non-commuting drift and diffusion matrices, a suitable choice of the weight functions is generally more challenging.

Our idea is the following: the asymptotic stability of the zero solution of the transformed system (14) can be determined by its Lyapunov exponent Λ represented by (11) where the kernel $\widetilde{\mathcal{K}}(\cdot)$ now takes the form

$$\widetilde{\mathcal{K}}(\lambda) = \left\langle \left(A^h - \sigma \sum_{i=1}^{N} d_i G_i^h \right) \lambda, \lambda \right\rangle + \frac{1}{2} \operatorname{trace}\left[D(\lambda)\right] - \langle D(\lambda)\lambda, \lambda \rangle .$$

As before, see Eq. (12) in Section 2.2, the definiteness of $\widetilde{\mathcal{K}}(\cdot)$ is determined by the definiteness of the matrix

$$\widetilde{K}(\lambda) := A^h - \sigma \sum_{i=1}^{N} d_i G_i^h - D(\lambda) + \frac{\sigma^2}{2} \sum_{k=1}^{N} \left(G_k^h\right)^2$$

on $\mathcal{S}^{N-1}$. In fact for $\widetilde{K}(\lambda)$ (strictly) positive or negative definite, the function $\widetilde{\mathcal{K}}(\lambda)$ is (strictly) positive or negative definite and in consequence the top Lyapunov exponent is positive or negative. As such, obtaining a positive top Lyapunov exponent is not desirable, as then almost all the trajectories of the transformed system (14) would explode, which would result in a wildly varying Monte Carlo estimator. However, the matrix $\widetilde{K}(\lambda)$ can also be indefinite, which essentially means $\widetilde{\mathcal{K}}(\lambda) \geq 0$ on some subset of $\mathcal{S}^{N-1}$, possibly yielding a larger Lyapunov exponent than in the case of strictly negative definite $\widetilde{K}(\lambda)$.

As the (negative) Lyapunov exponent Λ measures the rate of exponential decay of the trajectories of system (14), such an increase of Λ implies a slower decay of the solution trajectories. The main idea now is to establish exactly this situation with judicious choices of the d_i.

As the term $\sum_{i=1}^{N} d_i G_i^h$ is a diagonal matrix, we cannot modify off-diagonal entries. Therefore we use the weights d_i to compensate the influence of the diagonal entries of $-D(\lambda) + \frac{\sigma^2}{2} \sum_{k=1}^{N} \left(G_k^h\right)^2$ which are given by

$$\left(-D(\lambda) + \frac{\sigma^2}{2} \sum_{k=1}^{N} \left(G_k^h\right)^2 \right)_{ii} = \sigma^2 (\frac{1}{2} - \lambda_i^2) \left(\sum_{k=1}^{N} (G_{k,ii}^h)^2 \right).$$

Measuring the distance between the diagonals with a least-squares approach we define the optimisation problem

$$\underset{d_1,\ldots,d_N \in \mathbb{R}}{\operatorname{argmin}} \left[\operatorname{trace} \left[\left(\tilde{p} \sum_{i=1}^{N} \sigma^2 (G_i^h)^2 + \sigma \sum_{i=1}^{N} d_i G_i^h \right)^2 \right] \right]. \tag{17}$$

The parameter $\tilde{p} > 0$ allows us to control the definiteness of $\widetilde{\mathcal{K}}$, in principle one could choose $\widetilde{p}$ sufficiently large such that the Lyapunov exponent Λ gets positive.

The optimisation problem (17) requires us to solve a linear system of dimension N.

Alternatively we can decompose (17) into N optimisation problems of the form

$$\operatorname*{argmin}_{d_i\in\mathbb{R}}\left[\operatorname{trace}\left[\left(\tilde{p}\sigma^2(G_i^h)^2+\sigma d_i G_i^h\right)^2\right]\right]=-\tilde{p}\sigma\,\frac{\sum_{k=1}^N(G_{i,kk}^h)^3}{\sum_{k=1}^N(G_{i,kk}^h)^2} \tag{18}$$

for $1\leq i\leq N$, which can be solved analytically. For both cases the weights d_i with even indices i are essentially zero, whereas for odd indices the values of $|d_i|$ decrease with growing i, see Table 1 in Section 4 for the values d_i, where $N=10$. The corresponding density process can be calculated explicitly and is given by

$$\Theta(t)=\exp\left[\left(-\sum_{i=1}^N d_i^2\right)\frac{t}{2}+\sum_{i=1}^N d_i\beta_i(t)\right]. \tag{19}$$

Note that for the stability analysis of the discretised stochastic heat equation (8), we required that the noise is non-degenerate in the sense of Condition (10) to obtain the representation (11) for the top Lyapunov exponent. Therefore, the number of Wiener processes was required to be greater or equal N. However, the importance sampling technique proposed above (and thus the optimisation problems (17) and (18)) can also be modified for Eq. (8), when the number J of Wiener processes is smaller than the dimension N by appropriately truncating the corresponding sums after the first J elements.

4 Numerical results

4.1 Monte Carlo estimators

Using the measure transformations from Section 3.2, we define the Monte Carlo estimator

$$\widetilde{E}_N[f(\widehat{X}(T)]=\frac{1}{N}\sum_{i=1}^N f(\widehat{X}^{(i)}(T))\Theta^{(i)}(T), \tag{20}$$

where $\widehat{X}^{(i)}(T)$ are numerical trajectories of the transformed system (14) and $\Theta^{(i)}(T)$ are the corresponding trajectories of the density process given by (19). For calculating the numerical trajectories we use the Euler-Maruyama method with a suitable time step size for the geometric Brownian motion (3), whereas for the stochastic heat equation (6) we employ the stochastic trapezoidal rule

$$\widehat{X}(t_{n+1})=\widehat{X}(t_n)+\frac{\Delta t}{2}\widetilde{A}^h\widehat{X}(t_{n+1})+\frac{\Delta t}{2}\widetilde{A}^h\widehat{X}(t_n)+\sigma\sum_{r=1}^N G_r\widehat{X}_n\Delta\beta_r^n,$$

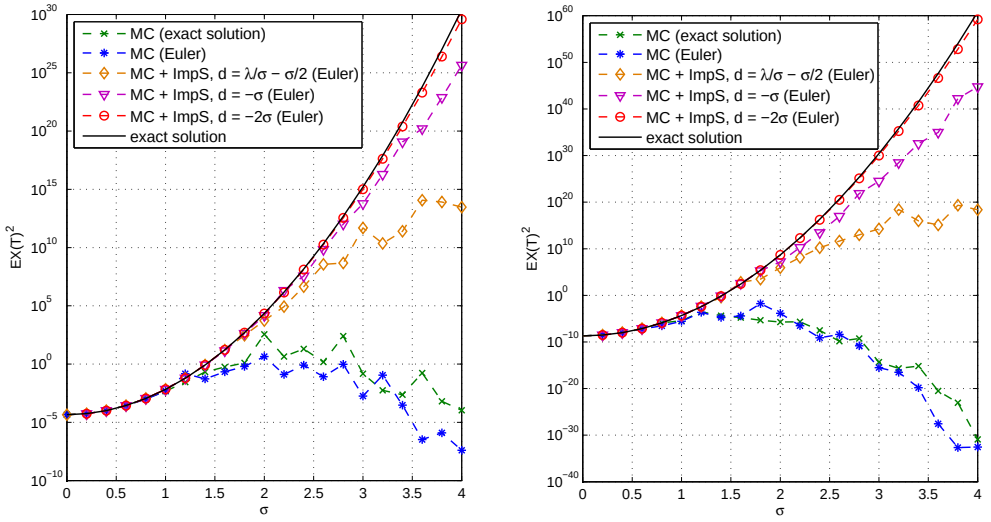
with $\beta_r^n\overset{iid}{\sim}\mathcal{N}(0,\sqrt{t_{n+1}-t_n})$, since the numerical treatment of the discretised stochastic heat equation (8) requires us to use a numerical method with suitable stability properties. It is straightforward to extend (20) to a multilevel Monte Carlo estimator which we denote by $\widetilde{E}^L[f(\widehat{X}(T)]$. We want to estimate the quantity $\mathbb{E}|X(T)|^2$ and compare our estimators with single and multilevel Monte Carlo estimators without modification.

A reference solution can be obtained by using results from Section 2.2. For this, let $\mathbb{E}Y(T) = (\mathbb{E}Y_1(T), \dots, \mathbb{E}Y_{N^2}(T))$ be the solution of the deterministic system (9) used for the mean-square stability analysis of the semi-discretised stochastic heat equation. Since $Y(t) = \mathrm{vec}(X(t)X(t)^T)$, we obtain a reference solution by

$$\mathbb{E}|X(T)|^2 = \sum_{i=1}^{N} \mathbb{E}X_i^2(T) = \sum_{i=0}^{N-1} \mathbb{E}Y_{(i\cdot N)+1}(T). \tag{21}$$

4.2 Geometric Brownian motion

In Figure 1, we compare the performance of the standard Monte Carlo estimator and the estimator defined in Eq. (20) applied to the geometric Brownian motion (3) with respect to different diffusion parameter σ. For these numerical experiments, we fix the drift parameter $\lambda = -1$ and the end time points $T = 5$ and $T = 10$. The time step size for the Euler-Maruyama scheme, $\Delta t = 2^{-8}$, is chosen such that the stability of the numerical method is guaranteed for all considered test cases. Furthermore, all Monte Carlo simulations are based on $M = 10^7$ independent realisations.



(a) $T = 5$ (b) $T = 10$

Figure 1: Monte Carlo simulations for the geometric Brownian motion (3) for $\lambda = -1$, and varying σ. All Monte Carlo simulations use $M = 10^7$ trajectories.

As we can see in Figure 1, the standard Monte Carlo estimator fails to reproduce the correct dynamics of the mean-square process. Note that there is hardly any difference in the qualitative behaviour of the standard Monte Carlo estimator using the exact solution or the numerical approximation based on the Euler-Maruyama scheme.

For the Girsanov transformation, we compare different choices of the parameter $\tilde{p}$. Here the solution of the optimisation problem (17) takes the form $d = -\sigma\tilde{p}$. First, the choice $\tilde{p} = -\frac{\lambda}{\sigma^2} + \frac{1}{2}$ (note that $-\lambda > 0$ and $d = \lambda/\sigma - \sigma/2$) eliminates the deterministic part in the

exact solution of the transformed system, i.e.

$$\widetilde{X}(t) = x_0 \exp\left(\sigma\beta(t)\right)$$

and for this reason, the zero solution is neither asymptotically a.s. stable nor unstable. The resulting estimation is already a substantial improvement compared to standard Monte Carlo estimators, however the mean-square process is still underestimated. Second, corresponding to our considerations in Section 3.2 increasing the parameter $\tilde{p}$ leads to a larger Lyapunov exponent. Since for large values of σ, it holds that $-\frac{\lambda}{\sigma^2} + \frac{1}{2} < 1$, we choose $d = -\sigma$, i.e. $\tilde{p} = 1$. For this choice the modified Monte Carlo estimator provides a better approximation of $\mathbb{E}\left[X^2(T)\right]$. Finally, the case $\tilde{p} = 2$ results in the optimal choice $d^* = -2\sigma$ (see Remark 3.1).

4.3 Stochastic heat equation

In Figure 2 and 3, we compare the qualitative behaviour of the standard and multilevel Monte Carlo estimators with and without measure transformation. We want to estimate the quantity $\mathbb{E}|X(T)|^2$ for different σ and time points T. For the Q-Wiener process, we choose the regularity parameter $r = 1$. To obtain reference solutions we carefully integrate the deterministic Equation (9) and use the representation (21). For the standard Monte Carlo simulations we use $M = 2^{11}$ realisations and a constant time step size $\Delta t = 2^{-11}$. For the multilevel Monte Carlo simulations we use 5 levels, where we calculate $M_\ell = 2^{11-\ell}$ trajectories with time step sizes $\Delta t_\ell = 2^{-(7+\ell)}$ on the corresponding refinement level $\ell = 1, \ldots, 5$. The Monte Carlo estimations after measure transformation (denoted by *MC + ImpS* and *MLMC + ImpS* in the plots) are computed with the same set of parameters. Depending on the optimisation criteria of Section 3.2 we distinguish between *OPT1*, where we minimised Eq. (17), and *OPT2*, where we used Eq. (18). In Table 1 below, the computed weights $d_i, i = 1, \ldots, N$, are given for $N = 10$ for both optimisation problems. In Figures 2 and 3 we see that the standard as well as the multilevel Monte Carlo estimator fail to approximate $\mathbb{E}|X(T)|^2$ correctly for $\sigma > 2$. Note that for $2 \leq \sigma \leq 3.4$ the zero solution of system (6) is still asymptotically mean-square stable, however the estimator substantially underestimates $\mathbb{E}|X(T)|^2$ due to the rapid decrease of the solution paths. For the estimators after measure transformation we see that the quality of the approximation strongly depends on the choice of the parameter $\tilde{p}$ as we have discussed in Section 3.2. For $\tilde{p} = 1/2$ the matrix $\widetilde{K}(\lambda)$ is still negative definite. The results are obviously better than for the untransformed estimators, nevertheless the approximations are not satisfying, especially for the larger time horizon $T = 5$. Increasing the parameter $(\tilde{p} = 1, \frac{3}{2})$ leads to an indefinite matrix $\widetilde{K}(\lambda)$. For both choices, we obtain good approximations of the qualitative behaviour of $\mathbb{E}|X(T)|^2$.

5 Conclusions

In this paper we analysed the Monte Carlo error for estimating the second moment of the solution process of a linear system of SODEs with asymptotically a.s. stable but mean-square

Table 1: List of computed non-zero weights $d_i, i = 1, 3, 5, 7, 9$, based on optimisation problems OPT1 (17) and OPT2 (18). All weights with even indices are essentially zero.

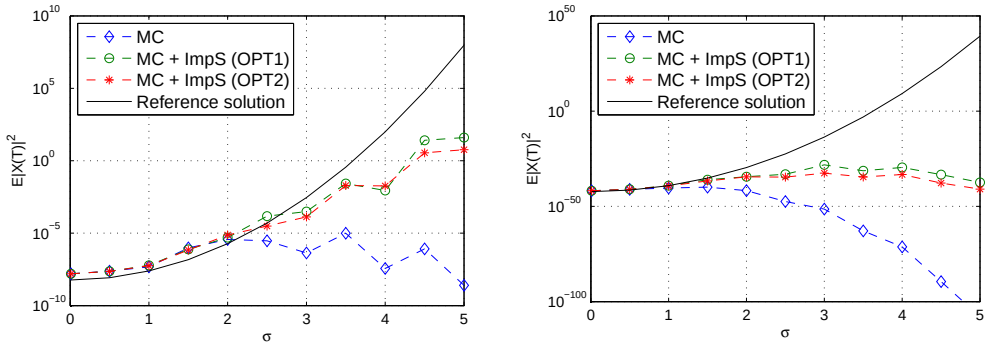
non-zero weights	OPT1	OPT2
d_1	$-1.201e-02$	$-1.385e-02$
d_3	$-3.877e-04$	$3.325e-03$
d_5	$-1.255e-04$	$2.316e-03$
d_7	$9.046e-05$	$1.469e-03$
d_9	$4.452e-07$	$5.832e-04$

unstable zero solution. The mean-square instability of the equilibrium of such systems is due to very rare, exploding trajectories. Since the standard Monte Carlo estimator essentially computes the average over a finite number of realisations, the path-wise behaviour of the numerical trajectories compromises the estimation and the Monte Carlo error for estimating $\mathbb{E}f(X(t))$, where $f(x)$ is a function with (at least) quadratic growth in $|x|$, is huge.

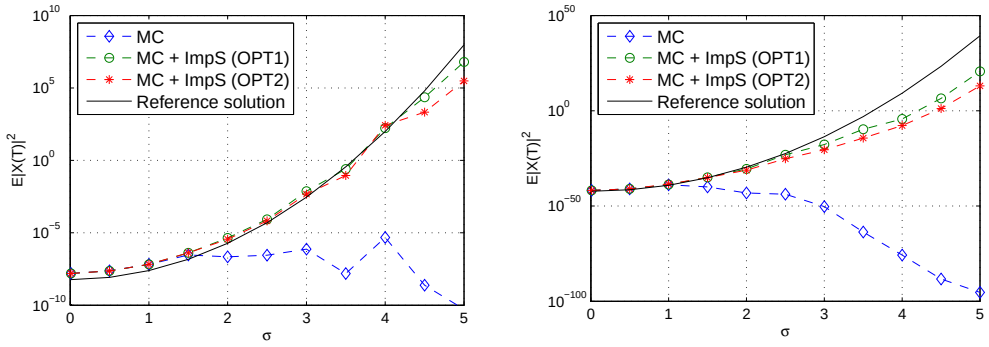
We further remark that this situation also causes difficulties in numerical experiments testing the weak convergence properties of numerical methods for SDEs. Such numerical illustrations can be performed by simulations of the quantity $|E_M[\widehat{X}^2(T)] - \mathbb{E}X^2(T)|$. However, for problems where the path-wise behaviour of the numerical realisation is totally different to the dynamics of the mean-square process, the Monte Carlo error dominates the systematic error by several orders of magnitude. Reducing the time step size will not lead to the expected error reduction in numerical simulations.

As we showed in Section 3.1, with the straightforward application of standard variance reduction techniques it is not possible to overcome these difficulties in our setting unless they focus on rare event simulation. For this reason, we proposed an appropriately tuned importance sampling technique which allows the trajectories to explode more frequently. We want to emphasise that the developed measure transformation does not change the linearity of the SDE system and can be precomputed. Thus, the proposed method does not increase the computational complexity of simulating a realisation of the quantity of interest. Furthermore, it is also possible to apply additional variance reduction methods to the modified estimator for further reducing its variance.

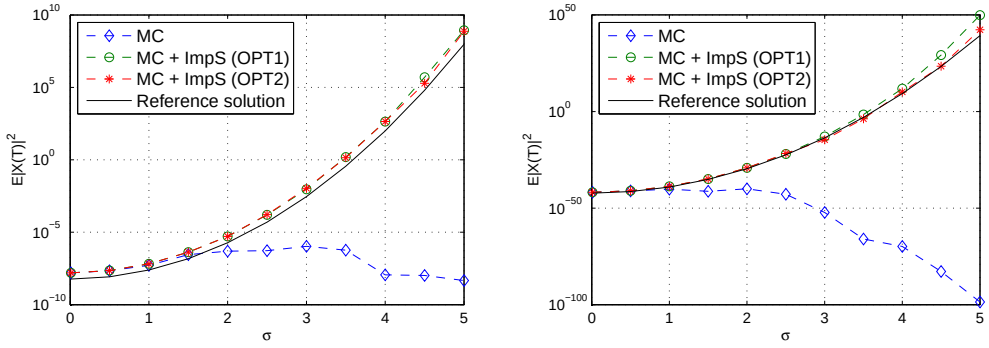
In this work we also compared standard and multilevel Monte Carlo methods with and without the proposed importance sampling technique and illustrated the improvement by the proposed method in numerical experiments for a spatially discretised stochastic heat equation.



(a) $T = 1$, $\tilde{p} = \frac{1}{2}$ (b) $T = 5$, $\tilde{p} = \frac{1}{2}$

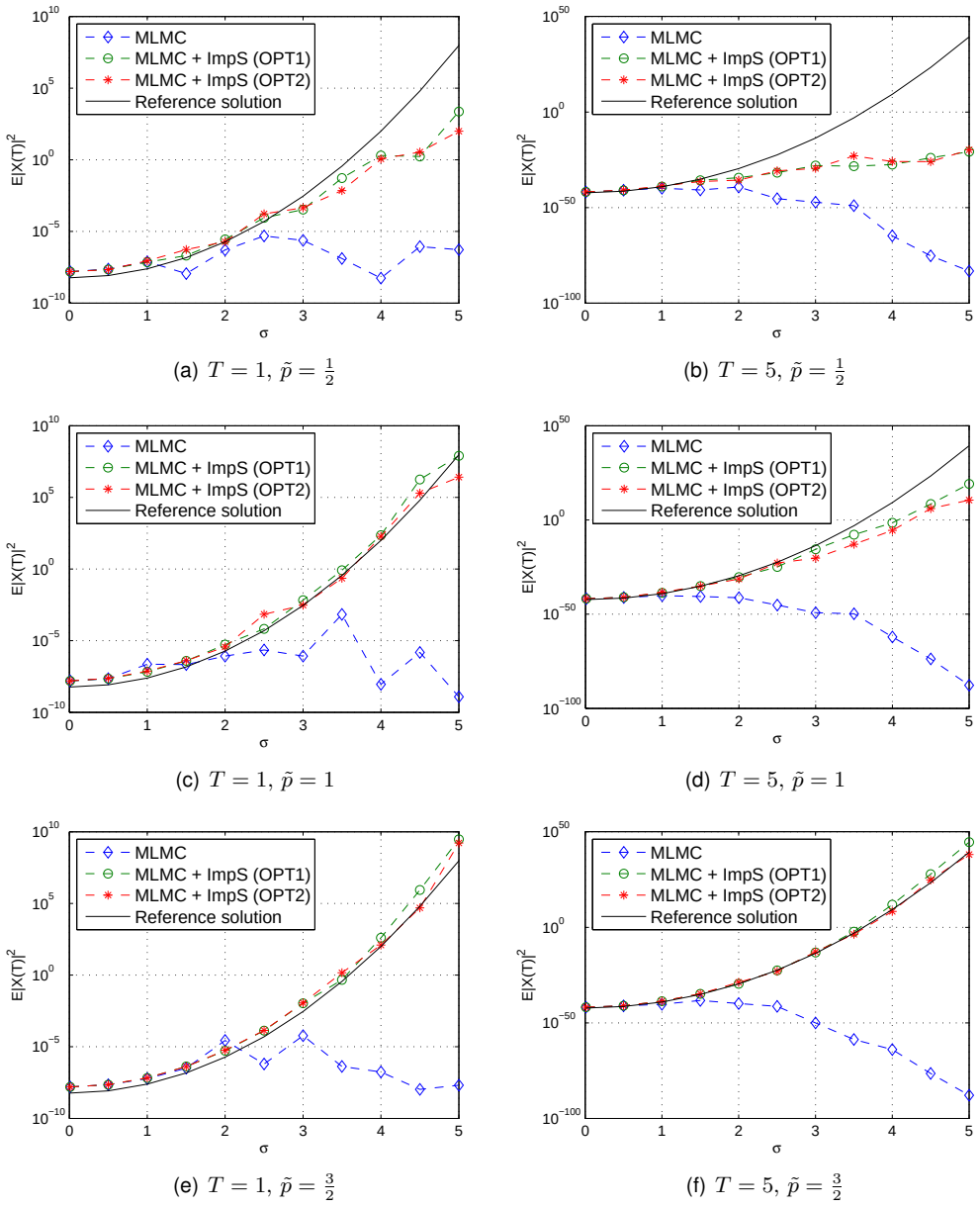


(c) $T = 1$, $\tilde{p} = 1$ (d) $T = 5$, $\tilde{p} = 1$



(e) $T = 1$, $\tilde{p} = \frac{3}{2}$ (f) $T = 5$, $\tilde{p} = \frac{3}{2}$

Figure 2: Monte Carlo methods for the spatially discretised stochastic heat equation (8) for $N = 10$ and varying σ. The reference solution is computed by Eq. (21).
Left column: Importance sampling with differently penalised weights at $T = 1$.
Right column: Importance sampling with differently penalised weights at $T = 5$.



(a) $T = 1$, $\tilde{p} = \frac{1}{2}$ (b) $T = 5$, $\tilde{p} = \frac{1}{2}$

(c) $T = 1$, $\tilde{p} = 1$ (d) $T = 5$, $\tilde{p} = 1$

(e) $T = 1$, $\tilde{p} = \frac{3}{2}$ (f) $T = 5$, $\tilde{p} = \frac{3}{2}$

Figure 3: Multilevel Monte Carlo methods for the spatially discretised stochastic heat equation (8) for $N = 10$ and varying σ. The reference solution is computed by Eq. (21).
Left column: Importance sampling with differently penalised weights at $T = 1$.
Right column: Importance sampling with differently penalised weights at $T = 5$.

References

[1] L. Arnold. “A formula connecting sample and moment stability of linear stochastic systems”. In: *SIAM J. Appl. Math.* 44.4 (1984), pp. 793–802.

[2] L. Arnold. *Stochastic differential equations: theory and applications*. Translated from the German. Wiley-Interscience [John Wiley & Sons], New York-London-Sydney, 1974.

[3] S. Asmussen and P.W. Glynn. *Stochastic Simulation. Algorithms and Analysis*. Stochastic Modelling and Applied Probability, 57. Springer, New York, 2007.

[4] E. Buckwar and T. Sickenberger. “A structural analysis of asymptotic mean-square stability for multi-dimensional linear stochastic differential systems”. In: *Applied Numerical Mathematics* 62.7 (2012), pp. 842–859.

[5] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and its Applications, 44. Cambridge University Press, Cambridge, 1992.

[6] M. B. Giles. “Multilevel Monte Carlo methods”. In: *Acta Numer.* 24 (2015), pp. 259–328.

[7] M. B. Giles. “Multilevel Monte Carlo path simulation”. In: *Oper. Res.* 56.3 (2008), pp. 607–617.

[8] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Vol. 53. Springer Science & Business Media, 2003.

[9] R. A. Horn and C. R. Johnson. *Matrix analysis*. Second Edition. Cambridge University Press, Cambridge, 2013.

[10] P. Imkeller and C. Lederer. “Some formulas for Lyapunov exponents and rotation numbers in two dimensions and the stability of the harmonic oscillator and the inverted pendulum”. In: *Dyn. Syst.* 16.1 (2001), pp. 29–61.

[11] A. Kebaier. “Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing”. In: *Ann. Appl. Probab.* 15.4 (2005), pp. 2681–2705.

[12] C. Kelly. “Stochastic stability analysis of a reduced galactic dynamo model with perturbed α-effect”. In: *Phys. A* 457 (2016), pp. 480–491.

[13] R. Khasminskii. *Stochastic Stability of Differential Equations*. Stochastic Modelling and Applied Probability, 66. Springer, Heidelberg, 2012.

[14] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Applications of Mathematics, 23. Springer-Verlag, Berlin, 1992.

[15] F. Kozin and S. Sugimoto. “Relations between sample and moment stability for linear stochastic differential equations”. In: *Proceedings of the Conference on Stochastic Differential Equations and Applications (Park City, Utah, 1976)*. Academic Press, New York, 1977, pp. 145–162.

[16] K. Liu. *Stability of Infinite Dimensional Stochastic Differential Equations with Applications*. Chapman & Hall/CRC Monographs, Surveys in Pure, and Applied Mathematics 135, 2006.

[17] G. J. Lord, C. E. Powell, and T. Shardlow. *An Introduction to Computational Stochastic PDEs*. Cambridge Texts in Applied Mathematics. Cambridge University Press, New York, 2014.

[18] X. Mao. *Stochastic Differential Equations and their Applications*. Horwood Publishing Series in Mathematics & Applications. Horwood Publishing Limited, Chichester, 1997.

[19] G. N. Milstein and M. V. Tretyakov. *Stochastic Numerics for Mathematical Physics*. Scientific Computation. Springer-Verlag, Berlin, 2004.

[20] P. E. Protter. *Stochastic integration and differential equations*. Second Edition. Vol. 21. Stochastic Modelling and Applied Probability. Springer-Verlag, Berlin, 2005.

[21] W.-H. Steeb and Y. Hardy. *Matrix calculus and Kronecker product. A practical approach to linear and multilinear algebra*. Second. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2011.

[22] D. Talay and L. Tubaro. “Expansion of the global error for numerical schemes solving stochastic differential equations”. In: *Stochastic Anal. Appl.* 8.4 (1990), pp. 483–509.

3 Importance sampling techniques for stochastic partial differential equations

E. Buckwar and A. Thalhammer

In this work we consider importance sampling techniques for stochastic partial differential equations (SPDE) based on an infinite dimensional version of the well-known Girsanov theorem. For this we develop construction techniques for measure transformations that can be used in numerical experiments for variance reduction purposes of the infinite dimensional equations. The key advantage of the proposed methods is that these techniques are independent of the (spatial) discretisation of the SPDE and that the difference in the computational effort between simulating the numerical trajectories for the standard Monte Carlo estimator and for the importance sampling methods is only the approximation of a linear, one-dimensional SODE. Besides the analysis of the infinite dimensional framework, various approximation and implementation issues are discussed. We conclude by presenting numerical experiments showing the effectiveness of the proposed techniques due to a remarkable reduction of the Monte Carlo error.

Key words. Variance reduction techniques, Importance sampling, Stochastic partial differential equations, Monte Carlo methods

AMS subject classifications. 60H15, 60H35, 35R60, 65C05, 62J10

1 Introduction

In many fields of science, stochastic partial differential equations (SPDEs) are used to model problems with uncertainties. Since the importance of such SPDE-based models increases constantly, it is necessary to develop efficient numerical methods in order to be able to cope with the computational complexity of the numerical experiments. In many of these applications, the quantities of interest are the statistics of the underlying solution process at a given time T, i.e. we want to estimate the quantity $\mathbb{E}[\varphi(X(T))]$ for a sufficiently smooth mapping $\varphi : H \to B$, where H and B are separable Hilbert spaces and $X(T)$ denotes the H-valued mild solution of a semilinear SPDE. Due to the fact that it is in general not possible to compute the expectation of $\varphi(X(T))$ explicitly, we have to rely on suitable estimators. Thus, besides issues concerning the spatial and temporal discretisation of the SPDE and the approximation of the noise, we have to additionally *discretise the probability space*, when we want to estimate the expectation.

The goal of this work is to derive and implement numerical methods that enhance the efficiency of the standard Monte Carlo estimator given by

$$\mathbb{E}[\varphi(X(T))] \approx E_M[\varphi(X(T))] := \frac{1}{M}\sum_{i=1}^{M}\varphi(X^{(i)}(T)),$$

where $X^{(i)}(T)$ are M independent realisations of the mild solution $X(T)$ defined in (2.4) below. For the resulting approximation error, also called the Monte Carlo error, the following result can be found in [4]: For a random variable $Y \in L^2(\Omega; B)$, i.e. $\mathbb{E}[\|Y\|_B^2] < \infty$, the Monte Carlo error can be expressed by

$$\|\mathbb{E}[Y] - E_M[Y]\|_{L^2(\Omega;B)} = \frac{1}{\sqrt{M}}\operatorname{Var}_B[Y]^{1/2}, \tag{1.1}$$

where the variance in the Hilbert space B is defined as $\operatorname{Var}_B[Y] := \mathbb{E}[\|Y - \mathbb{E}[Y]\|_B^2]$. Thus, increasing the number of samples M obviously leads to a decrease in the Monte Carlo error. However, this is not always suitable due to the high computational cost of the involved numerical realisations and the rather slow convergence rate of the Monte Carlo estimator (with respect to the number of samples M) attaining only $\mathcal{O}(M^{-1/2})$.

Especially, if we are dealing with approximations of mild solutions of SPDEs, an increase of the sample size is severely limited because of the complexity of the simulations. Thus, in order to make Monte Carlo estimators more efficient, we apply appropriate variance reduction techniques. These are Monte Carlo-type methods using instead of the original quantity of interest Y independent realisations of a random variable $\widetilde{Y}$ satisfying $\mathbb{E}[Y] = \mathbb{E}[\widetilde{Y}]$ and $\operatorname{Var}_B[\widetilde{Y}] \ll \operatorname{Var}_B[Y]$, see e.g. [17, 9] for an overview of variance reduction techniques in the finite dimensional setting. In this article, we develop measure transformations of the underlying probability measure $\mathbb{P}$ based on an infinite dimensional version of the Girsanov theorem in order to reduce the variance of the transformed random variables $\widetilde{Y}$. This variance reduction technique is called *importance sampling* and has already been successfully applied to SODEs, see [17, 15, 1].

Note that importance sampling in the context of variance reduction techniques for SPDEs has already been treated in the literature, see e.g. [1, 16]. There are also optimal measure transformations, where optimality has to be understood in the sense that the corresponding measure transformation leads to a vanishing Monte Carlo error, see e.g. Section 4.1 of this article and [16] for SPDEs driven by a finite number of standard Wiener processes. However, in both cases it is highly non-trivial to compute these optimal measure transformations or at least to find a sufficiently good approximation of them. Thus, such techniques are not optimal from a computational point of view.

In this work, we construct families of infinite dimensional measure transformations that can be used for variance reduction purposes in numerical simulations of $\mathbb{E}[\varphi(X(T))]$. These measure transformations are constructed in such a way that they can easily be implemented and that the computational effort to simulate the numerical trajectories of the transformed process is just slightly higher than the cost for simulating the original solution

trajectories. Theoretical optimality results such as the one presented in Section 4.1 are used as guidance for the design of such families of measure transformations and in particular motivate the use of path-dependent transformations. As numerical experiments in Section 6 show, the variance can be significantly reduced by using the considered transformations and thus, the Monte Carlo error based on the proposed importance sampling techniques is considerably reduced.

The authors in [1] presented an importance sampling technique based on finite dimensional considerations: First, the underlying SPDE (the one-dimensional stochastic heat equation) is spatially discretised by using finite differences on a fixed spatial mesh and afterwards importance sampling is applied to the resulting system of SODEs. In contrast, we apply the proposed importance sampling techniques directly to the infinite dimensional SPDE to reduce the variance of the original quantity of interest. Afterwards any spatial discretisation scheme can be applied to the transformed equation. This has the big advantage that the importance sampling method is independent of the chosen spatial discretisation scheme and consequently we do not have to exploit special structures of the finite dimensional SODE systems.

The flow chart in Figure 1 summarises the importance sampling techniques for SPDEs driven by infinitely many standard Wiener processes. In this work, we develop the approach following the bold arrows. The dashed arrows indicate alternative methods that have already been treated in the literature.

Besides importance sampling there are also other variance reduction techniques for SPDEs, e.g. control variates or multilevel Monte Carlo methods. Especially the latter methods, see e.g. [2, 4], reduce the complexity of estimating $\mathbb{E}[\varphi(X(T))]$ remarkably and we want to emphasize that these methods could be further enhanced by a coupling with the proposed importance sampling techniques from this article.

The outline of this paper is as follows: In Section 2 we briefly discuss the general framework and in Section 3, we recall basic results for measure transformations for semilinear SPDEs based on the infinite dimensional Girsanov theorem for Q-Wiener processes. In Section 3.1, we present a method how such measure transformations can be constructed. We are discussing in Section 4 how these measure transformations can be used in order to reduce the Monte Carlo error and we are able to prove the existence of an optimal measure transformation in Section 4.1, for which the Monte Carlo error vanishes completely. In Section 5, we provide a framework that enables us to couple the importance sampling techniques with spatial and temporal discretisation schemes. Finally, we conclude this article by presenting numerical experiments in Section 6, where we could achieve a remarkable improvement in the efficiency of the standard Monte Carlo estimator.

2 Framework

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$ denote a complete probability space with a filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$ satisfying the usual conditions. In this article we consider semilinear stochastic partial dif-

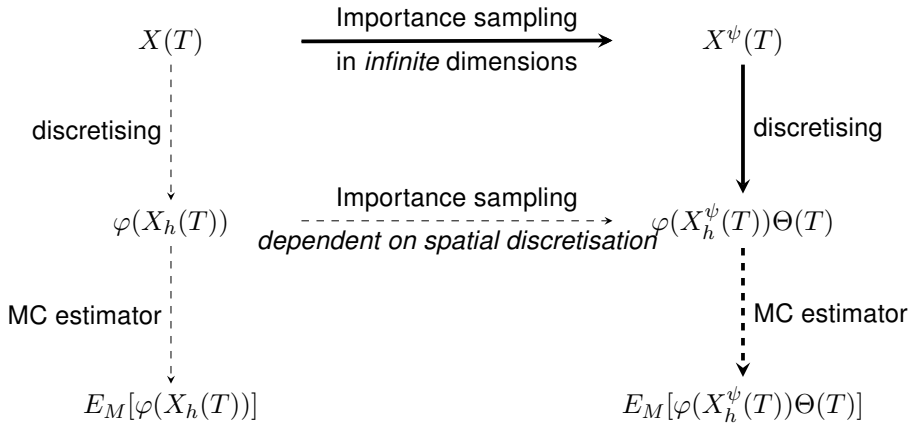


Figure 1: Overview of importance sampling techniques for stochastic partial differential equations driven by Q-Wiener processes.

ferential equations (SPDEs) on a separable Hilbert space $(H, \langle\cdot,\cdot\rangle_H)$ given by

$$\mathrm{d}X(t) = [AX(t) + F(X(t))]\,\mathrm{d}t + G(X(t))\,\mathrm{d}W(t), \qquad X(0) = X_0. \tag{2.1}$$

Here, $-A$ is assumed to be a densely defined, linear, symmetric and positive definite operator $-A : \mathcal{D}(-A) \subset H \to H$ and it is assumed to be the generator of an analytic semigroup $S(t), t \in [0,T]$. The domain of the fractional powers of the operator $-A$, which we denote by $\dot{H}^r := \mathcal{D}((-A)^{r/2})$, $r \in \mathbb{R}$, endowed with the inner product

$$\langle\cdot,\cdot\rangle_r := \langle(-A)^{r/2}\cdot, (-A)^{r/2}\cdot\rangle_H$$

also form separable Hilbert spaces, for details see [11] and the references therein. Using this notation, we assume that the initial value satisfies $X_0 \in L^2(\Omega, \dot{H}^1)$, where for any Hilbert space H

$$L^2(\Omega, H) := \{v : \Omega \to H \mid v \text{ is strongly measurable, } \|v\|^2_{L^2(\Omega;H)} := \mathbb{E}[\|v\|^2_H] < \infty\}.$$

Furthermore, let W be an H-valued Q-Wiener process with covariance operator $Q \in L(H)$ being a symmetric, non-negative definite, trace class operator. We denote by $\{e_k\}_{k\in\mathbb{N}}$ the eigenfunctions of Q, which form an orthonormal basis of H. Due to the trace class property of the covariance operator, the corresponding eigenvalues $\{\mu_k\}_{k\in\mathbb{N}}$ are summable, i.e. $\mathrm{tr}(Q) = \sum_{k=1}^{\infty}\mu_k < \infty$. Based on these properties the Q-Wiener process can be represented by (see e.g. [7, 19, 14] and the references therein)

$$W(t) = \sum_{k=1}^{\infty}\sqrt{\mu_k}e_k\beta_k(t), \tag{2.2}$$

where $\{\beta_k(t), t \in [0,T]\}_{k\in\mathbb{N}}$ is a sequence of independent, real-valued Brownian motions. Additionally, since Q is non-negative definite, the square-root of the operator Q is well-

defined, i.e. for all $\phi \in H$

$$Q^{1/2}\phi = \sum_{k=1}^{\infty} \sqrt{\mu_k}\langle \phi, e_k\rangle_H e_k \tag{2.3}$$

exists. In the following part, we denote by $Q^{-1/2}$ the pseudo-inverse of $Q^{1/2}$, see [19, Appendix C]. Using the representation of the square-root operator $Q^{1/2}$ given in Equation (2.3), we define the space $\mathcal{H}_0 := Q^{1/2}(H)$, which together with the norm $|\cdot|_0$ induced by the inner product

$$\langle \phi_1, \phi_2\rangle_0 = \langle Q^{-1/2}\phi_1, Q^{-1/2}\phi_2\rangle_H = \sum_{k=1}^{\infty} \frac{1}{\mu_k}\langle \phi_1, e_k\rangle_H \langle \phi_2, e_k\rangle_H, \qquad \phi_1, \phi_2 \in \mathcal{H}_0,$$

forms a separable Hilbert space, see [8], with orthonormal basis $\{\sqrt{\mu_k}e_k\}_{k\in\mathbb{N}}$.

Moreover, we denote by $L_{HS}(\mathcal{H}_0, H)$ the space of all Hilbert-Schmidt operators mapping from $\mathcal{H}_0$ to H. Finally, we assume that the drift operator $F : H \to H$ and the diffusion operator $G : H \to L_{HS}(\mathcal{H}_0, H)$ satisfy the following Lipschitz continuity and linear growth condition, see [4, Assumption 2.2]:

Assumption 2.1. *Assume that there exist constants $C_1, C_2 > 0$ such that for all $\phi, \phi_1, \phi_2 \in H$ it holds that*

$$\|F(\phi)\|_H + \|G(\phi)\|_{L_{HS}(\mathcal{H}_0,H)} \le C_1(1+\|\phi\|_H),$$
$$\|F(\phi_1) - F(\phi_2)\|_H + \|G(\phi_1) - G(\phi_2)\|_{L_{HS}(\mathcal{H}_0,H)} \le C_2\|\phi_1 - \phi_2\|_H.$$

If Assumption 2.1 is fulfilled, then Equation (2.1) has a unique H-valued mild solution $X(t)$ (see [11, Theorem 2.25]), i.e., there exists an up to modifications unique, $\mathcal{F}_t$-predictable, H-valued stochastic process $X : [0,T] \to H$ such that for all $t \in [0,T]$ it holds $\mathbb{P}$-a.s that

$$\sup_{t\in[0,T]} \|X(t)\|_{L^2(\Omega;H)} < \infty$$

and

$$X(t) = S(t)X_0 + \int_0^t S(t-s)F(X(s))\,\mathrm{d}s + \int_0^t S(t-s)G(X(s))\,\mathrm{d}W(s). \tag{2.4}$$

3 Measure transformations for SPDEs

In this section we first recall basic results on measure transformations for SPDEs. Afterwards we present in Section 3.1 a technique how these measure transformations can be constructed such that approximations of the involved transformation operators can be efficiently implemented for numerical simulations. Motivated by examples of finite dimensional measure transformations for SODE systems, we derive explicit representations of transfor-

mations in infinite dimensions, for which numerous numerical experiments are presented in Section 6.

The considered measure transformations are based on the following infinite dimensional version of the Girsanov theorem, see [7, Theorem 10.14]:

Theorem 3.1. *Assume that $\psi(t), t \in [0,T]$, is an $\mathcal{H}_0$-valued $\mathcal{F}_t$-predictable process such that the density process*

$$\Theta(t) = \exp\left(\int_0^t \langle \psi(s),\, \mathrm{d}W(s)\rangle_0 - \frac{1}{2}\int_0^t |\psi(s)|_0^2\,\mathrm{d}s\right)$$

is a martingale. Then, the process

$$\widehat{W}(t) = W(t) - \int_0^t \psi(s)\,\mathrm{d}s, \qquad t \in [0,T] \tag{3.1}$$

is a Q-Wiener process with respect to $\{\mathcal{F}_t\}_{t\in[0,T]}$ on the probability space $(\Omega, \mathcal{F}, \widehat{\mathbb{P}})$, where $\mathrm{d}\widehat{\mathbb{P}} = \Theta(T)\,\mathrm{d}\mathbb{P}$.

If the *weight operator* $\psi(t), t \in [0,T]$, is chosen appropriately according to Theorem 3.1, then one can apply the corresponding measure transformation to SPDE (2.1). Thus, we obtain with respect to the (transformed) probability space $(\Omega, \mathcal{F}, \widehat{\mathbb{P}})$ the following SPDE

$$\begin{aligned} X^\psi(t) - X_0 &= \int_0^t [AX^\psi(s) + F(X^\psi(s))]\,\mathrm{d}s + \int_0^t G(X^\psi(s))\,\mathrm{d}\widehat{W}(s) \\ &= \int_0^t [AX^\psi(s) + F(X^\psi(s)) - G(X^\psi(s))\psi(s)]\,\mathrm{d}s + \int_0^t G(X^\psi(s))\mathrm{dW(s)}. \end{aligned} \tag{3.2}$$

Since the coefficients F and G satisfy Assumption 2.1, there exists a unique mild solution $X^\psi(t)$ of the transformed SPDE (3.2) (with respect to the new probability measure $\widehat{\mathbb{P}}$) given by

$$\begin{aligned} X^\psi(t) = S(t)X_0 &+ \int_0^t S(t-s)[F(X^\psi(s)) - G(X^\psi(s))\psi(s)]\,\mathrm{d}s \\ &+ \int_0^t S(t-s)G(X^\psi(s))\,\mathrm{d}W(s). \end{aligned} \tag{3.3}$$

Note that $\Theta(t)$ can be represented in terms of the sequence of independent, scalar Brownian motions from the Karhunen-Loève expansion (2.2), i.e.

$$\Theta(t) = \exp\left(\sum_{k=1}^{\infty}\int_0^t \frac{1}{\sqrt{\mu_k}}\langle\psi(s), e_k\rangle_H\,\mathrm{d}\beta_k(s) - \frac{1}{2}\sum_{k=1}^{\infty}\int_0^t \frac{1}{\mu_k}\langle\psi(s), e_k\rangle_H^2\,\mathrm{d}s\right).$$

Thus, the density process $\Theta(t)$ can be seen as the solution of the one-dimensional SDE

driven by infinitely many independent one-dimensional Brownian motions, i.e.

$$\begin{aligned} \mathrm{d}\Theta(t) &= \Theta(t)\langle\psi(t),\, \mathrm{d}W(t)\rangle_0 = \Theta(t)\left(\sum_{k=1}^{\infty}\frac{1}{\sqrt{\mu_k}}\langle\psi(s), e_k\rangle_H\,\mathrm{d}\beta_k(t)\right), \\ \Theta(0) &= 1. \end{aligned} \tag{3.4}$$

Unfortunately, the martingale property of the density process $\Theta(t)$ in Theorem 3.1 is not fulfilled for arbitrary choices of $\psi(t)$ since in general $\Theta(t)$ is only a supermartingale and for this reason $\mathbb{E}[\Theta(T)] \leq 1$, see [13, Appendix I]. However, under the condition

$$\mathbb{P}\left(\int_0^T |\psi(s)|_0^2\,\mathrm{d}s < \infty\right) = 1 \tag{3.5}$$

the density process can be represented by a one-dimensional stochastic exponential, see [7, Lemma 10.15], given by

$$\Theta(t) = \exp\left(\int_0^t |\psi(s)|_0\,\mathrm{d}\beta(s) - \frac{1}{2}\int_0^t |\psi(s)|_0^2\,\mathrm{d}s\right) \quad \text{for } t \in [0,T], \tag{3.6}$$

where $\beta(t), t \in [0,T]$, denotes a scalar Brownian motion. If condition (3.5) is fulfilled, then it is sufficient to show that the stochastic exponential (3.6) is a martingale. Due to the fact that the considered stochastic process from Equation (3.6) is now one-dimensional and real-valued, standard results from finite dimensional stochastic calculus can be applied. Hence, the martingale property of (3.6) can be verified e.g. by Novikov's condition, see e.g. [18, 20],

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^t |\psi(s)|_0^2\,\mathrm{d}s\right)\right] < \infty. \tag{3.7}$$

3.1 Constructing measure transformations

In this section, we present a method to construct weight operators $\psi(t), t \in [0,T]$, that fulfil the rather strong assumptions of the Girsanov theorem for infinite dimensional Q-Wiener processes as stated in Theorem 3.1. Recall that a suitable weight operator ψ has to fulfil that it is an $\mathcal{H}_0$-valued and $\mathcal{F}_t$-predictable process and that the induced density process $\Theta(t)$ has to be a martingale. The proposed construction method looks as follows:

At first, we start by choosing an H-valued, $\mathcal{F}_t$-predictable process $\Psi(t)$ as an initial guess for the weight operator, where the particular choice of Ψ might be motivated by already existing finite dimensional examples. In the next step Ψ is projected onto $\mathcal{H}_0$ by using the mapping $P_{\mathcal{H}_0}$ given by

$$\psi(s) = P_{\mathcal{H}_0}\Psi(s) = \sum_{k=1}^{\infty}\mu_k\langle\Psi(s), e_k\rangle_H e_k \quad \text{for all } s \in [0,T], \tag{3.8}$$

which guarantees that $\psi(s) \in \mathcal{H}_0$ for all $s \in [0,T]$.

Since the density process can be represented by the stochastic exponential (3.6) under

Condition (3.5), we want to construct ψ in a way such that this condition is fulfilled. A sufficient condition for (3.5) to hold is that the weight operator is uniformly bounded in the $\mathcal{H}_0$-norm, i.e. there exists a constant $C_0 > 0$ such that $\sup_{s\in[0,T]} |\psi(s)|_0 \leq C_0 < \infty$. If the weight operator is uniformly bounded, then the martingale property of $\Theta(t)$ follows directly from Novikov's condition (3.7).

If we cannot show the uniform boundedness of the weight operator ψ (with respect to the $\mathcal{H}_0$-norm) we have to find a suitable truncation of the initial process Ψ denoted by $\overline{\Psi}$ such that there exists a constant $C_1 > 0$ with

$$\sup_{k\in\mathbb{N}} \sup_{s\in[0,T]} |\langle \overline{\Psi}(s), e_k\rangle_H| \leq C_1 < \infty.$$

Then, we obtain

$$|\psi(s)|_0^2 = \sum_{k=1}^{\infty} \mu_k \langle \overline{\Psi}(s), e_k\rangle_H^2 \leq C_1^2 \operatorname{tr}(Q) < \infty$$

and thus, Novikov's condition (3.7) is fulfilled.

In the remainder of this section, we present two examples of such measure transformations that are not only of theoretical interest, but can also been implemented for numerical experiments. For these two examples, we specify the Hilbert space H to be $L^2(D)$, where D denotes a bounded domain. Furthermore, we assume that the eigenfunctions $\{e_k\}_{k\in\mathbb{N}}$ of the covariance operator Q are uniformly bounded by a constant $C_e > 0$, i.e., there exists a constant $C_e > 0$ such that for all $j \in \mathbb{N}$ it holds that $\|e_j\|_{L^\infty(D)} \leq C_e < \infty$. An example of an SPDE that fulfils all these assumptions is given e.g. by the stochastic heat equation with multiplicative noise defined in Section 6.

Time-constant weight operator

The main idea of the time-constant weight operator is motivated by the importance sampling technique presented in [1]. For this let $\{\psi_k\}_{k\in\mathbb{N}}$ be a bounded, real-valued sequence, i.e., there exists a constant $M > 0$ such that $|\psi_k| \leq M$ for all $k \in \mathbb{N}$. Then, we define

$$\psi(s) = \sum_{k=1}^{\infty} \mu_k \psi_k \langle 1, e_k\rangle_H e_k \tag{3.9}$$

for all $s \in [0, T]$. Note that the weight operator (3.9) is uniformly bounded in the $\mathcal{H}_0$-norm, since it holds for all $k \in \mathbb{N}$ that $|\langle \psi_k, e_k\rangle_{L^2(D)}| \leq C_e M |D|$, where $|D|$ denotes the volume of domain D. For this reason, Condition (3.5) is automatically fulfilled and therefore, the corresponding density process

$$\Theta(t) = \exp\left(\sum_{k=1}^{\infty} \sqrt{\mu_k}\psi_k \langle 1, e_k\rangle_H \beta_k(t) - \frac{t}{2}\sum_{k=1}^{\infty} \mu_k \psi_k^2 \langle 1, e_k\rangle_H^2\right)$$

is a martingale by Novikov's condition (3.7). Thus, all conditions of Theorem 3.1 are fulfilled and the measure transformation based on the weight operator from Equation (3.9) can be applied.

Truncated path-dependent weight operator

We want to construct a weight operator $\psi : [0,T] \times H \to \mathcal{H}_0$ that also depends on the pathwise behaviour of the mild solution of SPDE (2.1). Note that such path-dependent weight operators and the corresponding measure transformations are commonly used for finite dimensional problems as it is described e.g. in [10, 16].

For constructing such path-dependent weight operators, we specify in this example that the diffusion operator is chosen to be a Nemytskii operator induced by a Lipschitz continuous function $\gamma : \mathbb{R} \to \mathbb{R}$, i.e., $(G(u)v)[x] := \gamma(u(x))v(x)$ for all $v \in \mathcal{H}_0$ and $x \in D$.

For the choice of the initial process Ψ, we consider a bounded approximation $\overline{G}$ of the diffusion operator G. Subsequently, $\overline{G}$ is based on a bounded Lipschitz continuous function $\overline{\gamma}$ that approximates (e.g. by truncation) the corresponding Lipschitz continuous function γ. More specifically, we assume that there exists a constant $C_G > 0$ such that $\overline{\gamma} : \mathbb{R} \to [-C_G, C_G]$. Note that for Nemytskii operators that are already induced by a bounded, Lipschitz continuous function γ, such as e.g. $\gamma(x) = \sin(x)$, $\overline{G}$ can be chosen as G. Examples of such truncated operators $\overline{G}$ are presented in Section 6.

Note that the operator $\overline{G}(\phi)$ is for all $\phi \in H$ a Hilbert-Schmidt operator mapping from $\mathcal{H}_0$ to H, since

$$\begin{aligned}\|\overline{G}(\phi)\|^2_{L_{HS}(\mathcal{H}_0,H)} &= \sum_{j=1}^{\infty} \|\overline{G}(\phi)\sqrt{\mu_j}e_j\|_H^2 = \sum_{j=1}^{\infty} \mu_j \int_D |\overline{\gamma}(\phi(x))e_j(x)|^2 \,\mathrm{d}x \\ &\leq C_G^2 \sum_{j=1}^{\infty} \mu_j \|e_j\|_H^2 = C_G^2 \,\mathrm{tr}(Q) < \infty.\end{aligned}$$

In the next step, we choose an element $\eta \in \mathcal{H}_0$ such that there exists a constant $C_\eta > 0$ with $\|\eta\|_H \leq C_\eta$ and define $\overline{\Psi}(s) := c\,\overline{G}(X^\psi(s))\eta$, where the scaling parameter $c \in \mathbb{R}$ controls the intensity of the measure transformation.

Since $\bar{\Psi}(s)$ is H-valued, we have to project $G(X^\psi(s))\eta$ onto $\mathcal{H}_0$. Thus, we define the truncated path-dependent weight operator as

$$\psi(s) = \psi(s, X^\psi(s)) := P_{\mathcal{H}_0}\overline{\Psi}(s) = c\sum_{k=1}^{\infty} \mu_k \langle \overline{G}(X^\psi(s))\eta, e_k\rangle_H e_k. \tag{3.10}$$

Note that instead of using a single scaling parameter c, one could also introduce a sequence $\{c_k\}_{k\in\mathbb{N}}$, where each of the elements c_k denotes a separate scaling for each of the summands in (3.10). However, for notational convenience we restrict ourselves to the

single parameter setting in this article. With this choice, we obtain

$$\begin{aligned}|\psi(s,X^{\psi}(s))|_0^2 &= \sum_{k=1}^{\infty}\frac{1}{\mu_k}\left\langle c\sum_{l=1}^{\infty}\mu_l\langle\overline{G}(X^{\psi}(s))\eta,e_l\rangle_H e_l,e_k\right\rangle_H^2\\ &= c^2\sum_{k=1}^{\infty}\mu_k\langle\overline{G}(X^{\psi}(s))\eta,e_k\rangle_H^2\\ &\leq (c\,C_G C_\eta)^2\,\mathrm{tr}(Q)<\infty.\end{aligned}$$

Hence, Conditions (3.5) and (3.7) are both fulfilled and for this reason the density process is a martingale. For the $\mathcal{F}_t$-predictability of the weight operator $\psi(t)$ it is sufficient to show that the transformed SPDE (3.2) has a unique mild solution $X^{\psi}(t)$ with respect to the transformed probability measure $\widehat{\mathbb{P}}$. However, this is automatically fulfilled since F and G satisfy Assumption 2.1 and for this reason, there exists a unique $\mathcal{F}_t$-predictable mild solution $X^{\psi}(t)$ to the transformed SPDE (3.2). Thus, all conditions from Theorem 3.1 are fulfilled and the measure transformation based on the path-dependent weight operator (3.10) can be applied.

4 Variance reduction by importance sampling

In this section, we examine how we can exploit the measure transformations introduced in Section 3 for variance reduction purposes. Based on the notation above, let $X(T)$ denote the (mild) solution of SPDE (2.1) and let $X^{\psi}(T)$ be the corresponding (mild) solution of the transformed SPDE (3.2) based on an appropriate weight operator ψ. Then, due to the Girsanov theorem, we get

$$\mathbb{E}_{\mathbb{P}}[\varphi(X(T))]=\mathbb{E}_{\widehat{\mathbb{P}}}[\varphi(X^{\psi}(T))]=\mathbb{E}_{\mathbb{P}}[\varphi(X^{\psi}(T))\Theta(T)],$$

where $\varphi : H\to B$ and $(B,\langle\cdot,\cdot\rangle_B)$ denotes a separable Hilbert space. Thus, by using a Monte Carlo estimator based on independent realisations of $\varphi(X^{\psi}(T))\Theta(T)$ instead of simulating independent samples of $\varphi(X(T))$ we obtain an unbiased estimator for $\mathbb{E}[\varphi(X(T))]$. The goal is now to find a weight operator $\psi(t), t\in[0,T]$, such that

$$\mathrm{Var}_B[\varphi(X^{\psi}(T))\Theta(T)]\ll\mathrm{Var}_B[\varphi(X(T))],$$

which by Equation (1.1) implies that the Monte Carlo error is reduced.

As it is also indicated in [9] for the finite dimensional setting, importance sampling is a strong tool for variance reduction and the choice of the weight operator $\psi(s), s\in[0,T]$, is crucial for a successful application of importance sampling. In particular, one has to be aware of the fact that it is also possible to enlarge the variance of the underlying random variable dramatically by choosing an inappropriate weight operator ψ.

From a computational point of view, it is also very important to find a balance between the computational cost of computing the importance sampling technique and its reduction

of the variance of the quantity of interest. Thus, optimal importance sampling techniques or at least approximations of such measure transformations might not be suitable for implementations if the computational cost for computing these measure transformations exceeds the computational complexity of solving the original problem.

For this reason, we proceed as follows: We first present in Section 4.1 an optimal measure transformation such that the variance of $\varphi(X^{\psi}(T))\Theta(T)$ vanishes. Afterwards, we discuss how we can use this optimal measure transformation as guidance for defining measure transformations that lead to variance reduction techniques that can be efficiently implemented. Numerical experiments in Section 6 show the effectiveness of the resulting importance sampling methods.

4.1 Optimal variance reduction

In this section, we want to optimise the above importance sampling technique with respect to the weight operator $\psi(t), t \in [0,T]$. The term *optimal* has to be understood in the sense that the variance of $\varphi(X^{\psi}(T))\Theta(T)$ vanishes for a fixed time point T with respect to the mapping φ. This is summarised in the following definition:

Definition 4.1. The importance sampling weight operator $\widetilde{\psi} : [0,T] \to U_0$ is called φ-*optimal* with respect to time T if

$$\mathrm{Var}_B[\varphi(X^{\widetilde{\psi}}(T))\Theta(T)] = 0.$$

Such optimal variance reduction techniques have already been developed for different types of SDEs. In [16, Chapter 4], the authors proposed an importance sampling technique combined with a control variate approach for strong solutions of parabolic SPDEs that would eliminate the Monte Carlo error completely. However, this is only possible if the SPDE (2.1) is driven by finitely many standard Wiener processes. Moreover, the resulting methods require the simulation of an additional SDE system and of additional Wiener processes.

In this section, we want to find such an optimal importance sampling method for mild solution of SPDEs driven by general Q-Wiener processes. The proof for this can be seen as the infinite dimensional counterpart of the proof presented in [17, Theorem 4.1], where the optimality result in finite dimensions is shown by using the solution of the corresponding Kolmogorov backward equation and the Itô formula. However, the techniques used in the proof of the finite dimensional result cannot be applied directly. As a remedy we construct the φ-optimal measure transformation as the limit of a sequence of optimal measure transformations, where the unbounded operator A is replaced by its Yosida approximations.

Remark 4.2. Results extending the Itô formula to mild solutions as presented in [6] cannot be applied in the same manner since the proof of the optimality result relies on a specific interplay of the solution of the underlying Kolmogorov backward equation and the Itô formula that is not present if one applies the mild Itô formula from [6].

For the optimal weight operator $\psi(t), t \in [0,T]$, we have to ensure additionally that there

exists a unique solution u to the Kolmogorov backward equation of SPDE (2.1) given by

$$\begin{aligned}\mathcal{L}u(t,x) = \frac{\partial u(t,x)}{\partial t} + \left\langle Ax + F(x), \frac{\partial u(t,x)}{\partial x} \right\rangle_H \\ + \frac{1}{2}\operatorname{tr}\left(\frac{\partial^2 u(t,x)}{\partial x^2}(G(x)Q^{1/2})(G(x)Q^{1/2})^* \right) = 0, \quad \text{for } t \in [0,T), \quad (4.1)\\ u(T,x) = \phi(x).\end{aligned}$$

For this reason, we have to introduce the following assumption within this section:

Assumption 4.3. *The first and second Fréchet derivatives of F and G are continuous and bounded and satisfy the two conditions:*

1 there exists a constant $M_1 > 0$ such that for all $x, y \in H$

$$\|DF(x)y\|_H + \|DG(x)y\|_{L_{HS}(\mathcal{H}_0,H)} \leq M_1\|y\|_H,$$

2 there exists a constant $M_2 > 0$ such that for all $x, y, z \in H$

$$\|D^2F(x)(y,z)\|_H + \|D^2G(x)(y,z)\|_{L_{HS}(\mathcal{H}_0,H)} \leq M_2\|y\|_H\|z\|_H.$$

Under this assumption there exists by [8, Theorem 3.11] a unique solution to (4.1) that is used to construct a φ-optimal weight operator $\psi(t), t \in [0,T]$ in the following theorem:

Theorem 4.4. *Let the initial value $X_0 \in \dot{H}^1$ be deterministic and let $\varphi : H \to \mathbb{R}_0^+$ be a bounded functional with bounded first and second Fréchet derivatives and $\varphi(h) > 0$ for all $h \in H \setminus \{0\}$. Additionally assume that the coefficients of SPDE* (2.1) *satisfy Assumptions 2.1 and 4.3. Finally let $u : [0,T] \times H \to \mathbb{R}$ denote the solution of the Kolmogorov backward equation* (4.1) *of SPDE* (2.1)*.*

Then, the choice

$$\begin{aligned}\widetilde{\psi}(s) &= \sum_{k=1}^{\infty} \widetilde{\psi}_k(s, X^{\widetilde{\psi}}(s))e_k \\ &= -\sum_{k=1}^{\infty} \frac{\mu_k}{u(s, X^{\widetilde{\psi}}(s))} \left\langle \frac{\partial u}{\partial x}(s, X^{\widetilde{\psi}}(s)), G(X^{\widetilde{\psi}}(s))e_k \right\rangle_H e_k\end{aligned} \quad (4.2)$$

is φ-optimal if the resulting density process $\Theta(t), t \in [0,T]$, is a martingale.

Proof. Let $\rho(A)$ be the resolvent set of A and let $0 < \ell_0 \in \rho(A)$. Then for $\ell \geq \ell_0$ we consider the SPDEs defined by

$$\begin{aligned}\mathrm{d}X_\ell^\psi(t) = \left[A_\ell X_\ell^\psi(t) + F(X_\ell^\psi(t)) - G(X_\ell^\psi(t))\psi_\ell(t)\right]\mathrm{d}t \\ + G(X_\ell^\psi(t))\,\mathrm{d}W(t),\end{aligned} \quad (4.3)$$

with initial value $X_\ell(0) = X_0$, where $A_\ell = \ell R(\ell; A)$ and $R(\ell; A)$ denotes the resolvent operator of A. By [12, Proposition 1.3.6], there exists a unique strong solution to SPDE (4.3)

and due to Assumptions 2.1 and 4.3, there exists a unique solution u_ℓ to the Kolmogorov Backward Equation given by (see [8, Theorem 3.11])

$$\begin{aligned}\mathcal{L}u_\ell(t,x) = \frac{\partial u_\ell(t,x)}{\partial t} + \left\langle A_\ell x + F(x), \frac{\partial u_\ell(t,x)}{\partial x} \right\rangle_H \\ + \frac{1}{2}\operatorname{tr}\left(\frac{\partial^2 u_\ell(t,x)}{\partial x^2}(G(x)Q^{1/2})(G(x)Q^{1/2})^*\right) = 0, \quad \text{for } t \in [0,T),\end{aligned} \tag{4.4}$$

$$u_\ell(T,x) = \phi(x).$$

Since there exists a unique strong solution to (4.3) we can apply the integration by parts formula from [5, Corollary 2.6.]

$$\begin{aligned}u_\ell(t, X_\ell^\psi(t))\Theta_\ell(t) = \; & u_\ell(0, X_0) \\ & + \int_0^t \left(\mathcal{L}u_\ell(s, X_\ell^\psi(s)) - \left\langle \frac{\partial u_\ell}{\partial x}(s, X_\ell^\psi(s)), G(X_\ell^\psi)(s)\psi_\ell(s)\right\rangle_H\right)\Theta_\ell(s)\,\mathrm{d}s \\ & + \int_0^t \sum_{n=1}^\infty \left\langle \frac{\partial u_\ell}{\partial x}(s, X_\ell^\psi(s)), G(X_\ell^\psi(s))Q^{1/2}e_n\right\rangle_H \Theta_\ell(s)\langle\psi_\ell(s), Q^{1/2}e_n\rangle_0\,\mathrm{d}s \\ & + \int_0^t u_\ell(s, X_\ell^\psi(s))\Theta_\ell(s)\langle\psi_\ell(s),\, \mathrm{d}W(s)\rangle_0 \\ & + \int_0^t \Theta_\ell(s)\left\langle \frac{\partial u_\ell}{\partial x}(s, X_\ell^\psi(s)), G(X_\ell^\psi(s))\,\mathrm{d}W(s)\right\rangle_H.\end{aligned} \tag{4.5}$$

Note that $\mathcal{L}u_\ell = 0$, since u_ℓ solves the Kolmogorov backward equation (4.4). In the next step, we show that the remaining deterministic integrals in Equation (4.5) are equal. For this, note that for all $s \in [0,T]$ and $n \in \mathbb{N}$

$$\langle\psi_\ell(s), Q^{1/2}e_n\rangle_0 = \frac{1}{\sqrt{\mu_n}}\langle\psi_\ell(s), e_n\rangle_H.$$

Hence, we get that

$$\begin{aligned}\int_0^t \sum_{n=1}^\infty \left\langle \frac{\partial u_\ell}{\partial x}(s, X_\ell^\psi(s)), G(X_\ell^\psi(s))Q^{1/2}e_n\right\rangle_H \Theta_\ell(s)\langle\psi_\ell(s), Q^{1/2}e_n\rangle_0\,\mathrm{d}s \\ = \int_0^t \sum_{n=1}^\infty \left\langle \frac{\partial u_\ell}{\partial x}(s, X_\ell^\psi(s)), G(X_\ell^\psi(s))e_n\right\rangle_H \Theta_\ell(s)\langle\psi_\ell(s), e_n\rangle_H\,\mathrm{d}s \\ = \int_0^t \left\langle \frac{\partial u_\ell}{\partial x}(s, X_\ell^\psi(s)), G(X_\ell^\psi(s))\sum_{n=1}^\infty\langle\psi_\ell(s), e_n\rangle_H e_n\right\rangle_H \Theta_\ell(s)\,\mathrm{d}s \\ = \int_0^t \left\langle \frac{\partial u_\ell}{\partial x}(s, X_\ell^\psi(s)), G(X_\ell^\psi(s))\psi_\ell(s)\right\rangle_H \Theta_\ell(s)\,\mathrm{d}s.\end{aligned}$$

Thus, the two deterministic integrals in Equation (4.5) are equal and therefore, the difference

between these two integrals vanishes. Consequently, Equation (4.5) reduces to

$$\begin{aligned} u_\ell(t, X_\ell^\psi(t))\Theta_\ell(t) &= u_\ell(0, X_0) + \int_0^t u_\ell(s, X_\ell^\psi(s))\Theta_\ell(s)\langle \psi_\ell(s),\, \mathrm{d}W(s)\rangle_0 \\ &\qquad + \int_0^t \Theta_\ell(s)\left\langle \frac{\partial u_\ell}{\partial x}(s, X_\ell^\psi(s)), G(X_\ell^\psi(s))\,\mathrm{d}W(s)\right\rangle_H \\ &=: I + II + III. \end{aligned}$$

Note that I is deterministic,

$$\begin{aligned} II &= \sum_{k=1}^\infty \frac{1}{\sqrt{\mu_k}} \int_0^t u_\ell(s, X_\ell^\psi(s))\Theta_\ell(s)\, \langle \psi_\ell(s), e_k\rangle_H \,\mathrm{d}\beta_k(s), \\ III &= \sum_{k=1}^\infty \int_0^t \Theta_\ell(s)\sqrt{\mu_k}\left\langle \frac{\partial u_\ell}{\partial x}(s, X_\ell^\psi(s)), G(X_\ell^\psi(s))e_k\right\rangle_H \mathrm{d}\beta_k(s). \end{aligned}$$

Thus, by defining for all $k \in \mathbb{N}$ and $s \in [0,T]$

$$\widetilde{\psi}_k^\ell(s, X_\ell^{\widetilde{\psi}_\ell}(s)) := -\frac{\mu_k}{u_\ell(s, X_\ell^{\widetilde{\psi}_\ell}(s))}\left\langle \frac{\partial u_\ell}{\partial x}(s, X_\ell^{\widetilde{\psi}_\ell}(s)), G(X_\ell^{\widetilde{\psi}_\ell}(s))e_k\right\rangle_H,$$

the importance sampling weight function

$$\widetilde{\psi}_\ell(s, X_\ell^{\widetilde{\psi}_\ell}(s)) = \sum_{k=1}^\infty \widetilde{\psi}_k^\ell(s, X_\ell^{\widetilde{\psi}_\ell}(s))e_k$$

is φ-optimal at time $t = T$ since for all $t \in [0,T]$ it holds that

$$u_\ell(t, X_\ell^{\widetilde{\psi}_\ell}(t))\Theta_\ell(t) = u_\ell(0, X_0)$$

is deterministic. This is in particular valid for time point $t = T$, at which by Equation (4.4) $u_\ell(T, X_\ell^{\widetilde{\psi}_\ell}(T))\Theta_\ell(T)$ coincides with $\varphi(X_\ell^{\widetilde{\psi}_\ell}(T))\Theta_\ell(T)$. Hence, by considering the limit with respect to $\ell \to \infty$ we get

$$\lim_{\ell\to\infty} \varphi(X_\ell^{\widetilde{\psi}_\ell}(T))\Theta_\ell(T) = \lim_{\ell\to\infty} u_\ell(0, X_0) = u(0, X_0),$$

where u denotes the unique solution of (4.1). Thus, $\mathrm{Var}[\varphi(X^{\widetilde{\psi}}(T))\Theta(T)] = 0$, where $\widetilde{\psi}$ is defined as in Equation (4.2). □

This optimal measure transformation can now be used as guidance to adjust the proposed measure transformations from Section 3.1 such that the used importance sampling technique results in a significant variance reduction. The path-dependent structure of the optimal weight operator $\widetilde{\psi}$ particularly motivates the use of the path-dependent weight operator defined in Equation (3.10) since this can be interpreted as a crude approximation of the optimal choice given by (4.2). In numerical experiments in Section 6, we examine how to choose the parameter c of the weight operator (3.10) in order to achieve the best possible

variance reduction.

5 Approximation of measure transformations

In this section, we recall approximation properties of spatial and temporal discretisation techniques and discuss how the importance sampling methods based on the measure transformations form Section 3.1 can be implemented for numerical experiments. The advantage of the proposed methods is that the computational complexity of simulating the transformed random variables and the simulation of the original mild solution process is of the same order. This is due to the fact that for the final implementation the computation of the density process $\Theta(t)$ reduces for the considered measure transformations to the approximation of the solution of a one-dimensional SODE driven by finitely many standard Wiener processes. Hence, the difference in the computational cost of the standard approach compared to the proposed importance sampling methods is limited to the (numerical) approximation of a one-dimensional linear SODE.

Since an explicit representation of the mild solution of an SPDE is not known in general, we have to additionally approximate its mild solution $X(T)$ by a (spatially and temporally) discretised approximation $X_{h,K}$, see e.g. [14, 11] for details on strong and weak numerical approximations of stochastic equations in infinite dimensions. By using this additional approximation, the overall error of the Monte-Carlo method can be estimated by, see e.g. [17],

$$\begin{aligned}
&\|\mathbb{E}[\varphi(X(T))] - E_M[\varphi(X_{h,K})]\|_{L^2(\Omega;B)} \\
&\quad \leq \underbrace{\|\mathbb{E}[\varphi(X(T))] - \mathbb{E}[\varphi(X_{h,K})]\|_B}_{\text{systematic error}} + \underbrace{\|\mathbb{E}[\varphi(X_{h,K})] - E_M[\varphi(X_h(T))]\|_{L^2(\Omega;B)}}_{\text{Monte Carlo error}} \\
&\quad = \|\mathbb{E}[\varphi(X(T))] - \mathbb{E}[\varphi(X_{h,K})]\|_B + \frac{\mathrm{Var}_B[\varphi(X_{h,K})]^{1/2}}{\sqrt{M}}.
\end{aligned}$$

The following lemma shows that the variance of the approximated mild solution (2.4) can be estimated by the strong error and the variance of the original quantity of interest. For the analysis of $\mathrm{Var}[\varphi(X_{h,K})]$, we assume that the mapping $\varphi : H \to B$ satisfies a Lipschitz continuity condition, i.e. there exists a constant $C_L > 0$ such that for all $\phi_1, \phi_2 \in H$ it holds that $\|\varphi(\phi_1) - \varphi(\phi_2)\|_B \leq C_L \|\phi_1 - \phi_2\|$.

Lemma 5.1. *Let $\varphi : H \to B$ be Lipschitz continuous. Then, there exists a constant $C > 0$ independent of h such that*

$$\mathrm{Var}_B[\varphi(X_{h,K})] \leq C\|X(T) - X_{h,K}\|^2_{L^2(\Omega;H)} + 3\,\mathrm{Var}_B[\varphi(X(T))].$$

Proof. By using the triangle inequality and the Lipschitz continuity of φ, we get

$$\begin{aligned}
\mathrm{Var}_B[\varphi(X_{h,K})] &= \mathbb{E}[\|\varphi(X_{h,K}) - \mathbb{E}[\varphi(X_{h,K})]\|_B^2] \\
&= \mathbb{E}[\|\varphi(X_{h,K}) \pm \varphi(X_{h,K}) \pm \mathbb{E}[\varphi(X(T))] - \mathbb{E}[\varphi(X_{h,K})]\|_B^2] \\
&\leq 3\Big(\mathbb{E}[\|\varphi(X(T)) - \varphi(X_{h,K})\|_B^2] + \|\mathbb{E}[\varphi(X(T))] - \mathbb{E}[\varphi(X_{h,K})]\|_B^2 \\
&\qquad + \mathbb{E}[\|\varphi(X(T)) - \mathbb{E}[\varphi(X(T))]\|_B^2]\Big) \\
&\leq C\|X(T) - X_{h,K}\|_{L^2(\Omega;H)}^2 + 3\,\mathrm{Var}_B[\varphi(X(T))].
\end{aligned}$$

Thus, the statement is shown. □

In the following two sections, we derive importance sampling techniques based on the infinite dimensional measure transformations from Section 3.1 combined with different spatial and temporal discretisation techniques, i.e. for the space discretisation we consider a finite difference method (in Section 5.1) and a Galerkin finite element method (in Section 5.2). For both spatial discretisation methods, we consider the backward Euler scheme for the time integration.

5.1 Approximation of the mild solution by finite differences

In this section, we consider a finite difference method for approximating the mild solution of the transformed SPDE (3.2). For simplicity, we consider a one-dimensional setting on a bounded domain, where we fix a spatial mesh $\mathcal{T}_h$ with $N = 1/h, h > 0$, spatial grid points, i.e.

$$\mathcal{T}_h := \{x_0 < x_1 < \cdots < x_N\}.$$

The operator A is assumed to be a differential operator and its approximation $\mathbf{A}_h \in \mathbb{R}^{(N+1)\times(N+1)}$ is computed by using discrete difference quotients on $\mathcal{T}_h$. Then, we obtain for the finite difference approximation $\mathbf{x}(t) = [X(t,x_0), \dots X(t,x_N)]^T$ the following finite dimensional SODE system

$$\begin{aligned}
\mathrm{d}\mathbf{x}(t) &= [\mathbf{A}_h\mathbf{x}(t) + \mathbf{F}(\mathbf{x}(t)) - \mathbf{G}(\mathbf{x}(t))\psi(t)]\,\mathrm{d}t + \mathbf{G}(\mathbf{x}(t))d\mathbf{W}(t) \quad \text{for } t \in [0,T], \\
\mathbf{x}(0) &= [X_0(x_0), \dots, X_0(x_N)]^T,
\end{aligned}$$

where $\mathbf{W}(t) = [W(t,x_0), \dots, W(t,x_N)]^T$, $\mathbf{F}(\mathbf{x}(t)) = [F(X(t,x_0)), \dots, F(X(t,x_N))]^T$, and $\mathbf{G}(\mathbf{x}(t))\psi(t) = [G(X(t,x_0))\psi(t,x_0), \dots, G(X(t,x_N))\psi(t,x_N)]^T$. The resulting SODE system can be approximated by using any appropriate time integration method. For this, we fix a partition θ_K of the time interval $[0,T]$ consisting of $(K+1)$ time points with not

necessarily equidistant time step sizes $\Delta t_j = t_j - t_{j-1}$, i.e.

$$\theta_K := \{0 = t_0 < t_1 < \cdots < t_K = T\}. \tag{5.1}$$

As an example of a time integration method we consider the backward Euler scheme which is given by $\mathbf{x}_0 = \mathbf{x}(0)$ and

$$\mathbf{x}_j = \mathbf{x}_{j-1} + \Delta t_j(\mathbf{A}_h\mathbf{x}_j + \mathbf{F}(\mathbf{x}_{j-1}) + \mathbf{G}(\mathbf{x}_{j-1})\psi(t_{j-1})) + \mathbf{G}(\mathbf{x}_{j-1})\Delta\mathbf{W}_j$$

for $j = 1, \ldots, K$, where $\mathbf{x}_j$ is an approximation of $[X(t_j, x_0), \ldots, X(t_j, x_N)]^T$ and $\Delta\mathbf{W}_j = \mathbf{W}(t_j) - \mathbf{W}(t_{j-1})$.

As already mentioned in the introduction, the authors in [1] already considered the coupling of finite difference approximations of SPDEs and importance sampling. At first, they fixed the spatial discretisation of the linear, one-dimensional stochastic heat equation (see Equation (6.1) below) based on finite differences with N spatial grid points and afterwards an importance sampling technique was applied to the resulting finite dimensional SODE system. However, the approach in [1] is equivalent to the truncated infinite dimensional importance sampling technique from above by using the time-constant weight operator from Equation (3.9). For this let the weight operator $\psi^J(s), s \in [0, T]$, be defined for all $s \in [0, T]$ by

$$\psi^J(s) := P_J(\psi(s)) = \sum_{k=1}^{J} \mu_k\psi_k\langle 1, e_k\rangle e_k,$$

where $J \in \mathbb{N}$ is sufficiently large and P_J denotes the orthogonal projector onto the first J eigenfunction of Q. Thus, for numerical experiments concerning finite difference approximations of SPDE (2.1) and importance sampling based on the weight operator from Section 3.1 we refer to [1].

5.2 Approximation of the mild solution by Galerkin methods

For simulations of the mild solution (2.4), we now use an approximation based on a Galerkin finite element method in space. For this, we choose a nested sequence $\mathcal{V} = (V_h)_{h\in(0,1]}$ of finite dimensional subspaces of H satisfying $V_h \subset \dot{H}^1 \subset H$.

Following the approach from [11], we consider the Ritz projection $R_h : \dot{H}^1 \to V_h$ and the orthogonal projection $P_h : H \to V_h$ defined by

$$\begin{aligned} \langle R_h v, v_h\rangle_1 &= \langle v, v_h\rangle_1 \qquad \text{for all } v \in \dot{H}^1, v_h \in V_h, \\ \langle P_h v, v_h\rangle_H &= \langle v, v_h\rangle_H \qquad \text{for all } v \in H, v_h \in V_h \end{aligned}$$

For the convergence analysis of the Galerkin finite element method we need the following assumption on the orthogonal projectors R_h and P_h:

Assumption 5.2. *For the family of finite dimensional subspaces $\mathcal{V} = (V_h)_{h\in(0,1]}$, there*

exists a constant $C > 0$ *such that for all refinement parameters* $h \in (0, 1]$ *it holds that*

I. $\|P_h v\|_1 \leq C\|v\|_1$ *for all* $v \in \dot{H}^1$,

II. $\|R_h v - v\|_H \leq Ch^s \|v\|_s$ *for all* $v \in \dot{H}^s$ *with* $s \in \{1, 2\}$.

According to the considered finite dimensional subspace V_h, we define the discrete operator A_h as the unique operator satisfying

$$\langle -Av_h, w_h\rangle_H = \langle v_h, w_h\rangle_1 = \langle -A_h v_h, w_h\rangle_H$$

for all $v_h, w_h \in V_h$. Note that $-A_h$ is a symmetric and positive definite operator on V_h and thus it is also the generator of an analytic semigroup. For this reason, there exists a unique mild solution $X_h(t)$ of the spatially discretised SPDE (2.1) given by

$$\begin{aligned} X_h(t) = S_h(t)P_h X_0 &+ \int_0^t S_h(t-s)P_h F(X_h(s))\,\mathrm{d}s \\ &+ \int_0^t S_h(t-s)P_h G(X_h(s))\,\mathrm{d}W(s). \end{aligned} \tag{5.2}$$

For a given weight operator $\psi(s), s \in [0, T]$, satisfying the conditions of Theorem 3.1, we can rewrite the transformed mild solution of the semi-discrete problem (5.2) as

$$\begin{aligned} X_h^\psi(t) &= S_h(t)P_h X_0 + \int_0^t S_h(t-s)P_h F(X_h^\psi(s))\,\mathrm{d}s \\ &\quad + \int_0^t S_h(t-s)P_h G(X_h^\psi(s))\,\mathrm{d}\widehat{W}(s) \\ &= S_h(t)P_h X_0 + \int_0^t S_h(t-s)P_h[F(X_h^\psi(s)) - G(X_h^\psi(s))\psi(s)]\,\mathrm{d}s \\ &\quad + \int_0^t S_h(t-s)P_h G(X_h^\psi(s))\,\mathrm{d}W(s). \end{aligned}$$

Thus, by using the density process $\Theta(t), t \in [0, T]$, we can also deduce for the semi-discrete approximation that

$$\mathbb{E}[\varphi(X_h(T))] = \mathbb{E}_{\widehat{\mathbb{P}}}[\varphi(X_h^\psi(T))] = \mathbb{E}[\varphi(X_h^\psi(T))\Theta(T)]. \tag{5.3}$$

For a fully discrete approximation of the mild solution we have to additionally consider a discretisation in time. For this, we choose again the backward Euler scheme on the time grid θ_K given in Equation (5.1)

$$\begin{aligned} X_{h,j} &= X_{h,j-1} + \Delta t_j\, A_h X_{h,j} + \Delta t_j P_h F(X_{h,j-1}) - P_h G(X_{h,j-1})\Delta W^j \quad j = 1, \ldots, K, \\ X_h^0 &= P_h X_0, \end{aligned}$$

where $\Delta t_j = t_j - t_{j-1}$ and $\Delta W^j = W(t_j) - W(t_{j-1})$ are the Wiener increments. Now, let $\Delta t = \max_i \Delta t_i$. Then, by Assumption 2.1 and 5.2, it follows that there exists a constant

independent of h and Δt such that, see [11, Theorem 3.14],

$$\|X(T) - X_{h,K}\|_{L^2(\Omega;H)} \leq C(h + \Delta t^{1/2}). \tag{5.4}$$

The following lemma provides an estimate of the overall error of the standard Monte Carlo estimator in terms of spatial and temporal refinement parameters and the variance of the original quantity of interest.

Lemma 5.3. *Let $\varphi : H \to B$ be Lipschitz continuous. Then, there exists a constant $C > 0$ independent of h and Δt such that*

$$\|\mathbb{E}[\varphi(X(T))] - E_M[\varphi(X_{h,K})]\|_{L^2(\Omega;B)} \leq C(h + \Delta t^{1/2}) + \frac{\sqrt{3}\,\mathrm{Var}_B[\varphi(X(T))]^{1/2}}{\sqrt{M}}.$$

Proof. Note that

$$\|\mathbb{E}[\varphi(X(T))] - E_M[\varphi(X_{h,K})]\|_{L^2(\Omega;B)} \leq \\ \|\mathbb{E}[\varphi(X(T))] - \mathbb{E}[\varphi(X_{h,K})]\|_B + \frac{1}{\sqrt{M}}\,\mathrm{Var}_B[\varphi(X_{h,K})]^{1/2}.$$

Since φ is Lipschitz, the first summand can be estimated by the strong approximation error from (5.4). Finally, by applying Lemma 5.1 and using again the strong convergence properties from (5.4), the statement is shown. □

Note that in cases where the systematic error is dominated by the Monte Carlo error a further reduction of the spatial and temporal grid width does not lead to an improvement in the error bound of Lemma 5.3. Thus, we propose to enhance the performance of the Monte Carlo simulation by using importance sampling, for which we show numerous numerical experiments in Section 6 using the path-dependent weight operator from Section 3.1.

Until now, we have only discussed the discretisation of SPDE (2.1) with respect to space and time. For numerical experiments, we have to consider in addition an appropriate noise approximation, where we truncate the Karhunen-Loève expansion by using

$$G^{J_0}(u) = G(u)P_{J_0}, \tag{5.5}$$

where P_J denotes the orthogonal projector onto $\mathrm{span}(e_1, \ldots, e_{J_0})$. Note that in order to preserve the convergence rates of the backward Euler scheme, we have to choose the truncation parameter $J_0 \in \mathbb{N}$ carefully, see e.g. [14, Lemma 10.33] and [3, Lemma 3.1].

5.3 Approximation of the density process

In this section, we discuss how to approximate the density process $\Theta(t), t \in [0, T]$, derived in Section 3.1. By the construction of the corresponding weight operators $\psi(t), t \in [0, T]$,

the resulting density processes are the strong solutions of SODEs of the following type:

$$\mathrm{d}\Theta(t) = \Theta(t)\left(\sum_{k=1}^{\infty}\sqrt{\mu_k}\,\kappa_k(t)\,\mathrm{d}\beta_k(t)\right), \qquad \Theta(0) = 1, \tag{5.6}$$

where the kernel functions $\kappa_k(s), s \in [0,T]$, are bounded, i.e. for all $s \in [0,T]$ and $k \in \mathbb{N}$ there exists a constant $C > 0$ such that $\kappa_k(s) \leq C < \infty$. Since in general we cannot compute the infinite series appearing in the density process explicitly, we introduce a truncation of the corresponding series that preserves the convergence properties of the used numerical method. Thus, we look for an appropriate truncation index $J_1 \in \mathbb{N}$ such that the truncated version of the SDE (5.6), i.e. for $t \in [0,T]$ let

$$\mathrm{d}\Theta^{J_1}(t) = \Theta^{J_1}(t)\left(\sum_{k=1}^{J_1}\sqrt{\mu_k}\,\kappa_k(t)\,\mathrm{d}\beta_k(t)\right), \qquad \Theta^{J_1}(0) = 1, \tag{5.7}$$

reduces to a one-dimensional linear SDE with autonomous coefficients (for the transformation based on the time-independent weight operator from Equation (3.9)) or with non-autonomous coefficients (for the path-dependent choice of $\psi(t), t \in [0,T]$ from Equation (3.10)).

The goal is to choose the truncation parameter J_1 in such a way that the corresponding weak convergence rates of the involved numerical method are preserved. In the following lemma we discuss conditions on how to choose J_1 for an arbitrary numerical method approximating the mild solution of SPDE (2.1) with given weak convergence order.

Lemma 5.4. *Assume that for all $t \in [0,T]$ there exists a constant $C > 0$ such that*

$$\|\varphi(X_h^{\psi}(t))\|_{L^2(\Omega;\mathbb{R})} \leq C < \infty$$

and that $\varphi : H \to \mathbb{R}$ is sufficiently smooth (according to the considered weak convergence order result). Furthermore, the numerical scheme approximating SPDE (2.1) *has given weak convergence order $\mathcal{O}(h^{\alpha})$ (under the assumption that the time step size Δt and the spatial refinement parameter h are appropriately coupled). Then it holds that the weak convergence order is preserved if $J_1 \in \mathbb{N}$ is chosen such that*

$$\sum_{k=J_1+1}^{\infty}\mu_k = \mathcal{O}(h^{2\alpha}).$$

Proof. First note that

$$\begin{aligned}
&|\mathbb{E}[\varphi(X(T))] - \mathbb{E}[\varphi(X_h^{\psi}(T))\Theta^{J_1}(T)]| \\
&\qquad = |\mathbb{E}[\varphi(X(T))] - \mathbb{E}[\varphi(X_h^{\psi}(T))\Theta^{J_1}(T)] \pm \mathbb{E}[\varphi(X_h(T))]| \\
&\qquad \leq |\mathbb{E}[\varphi(X(T))] - \mathbb{E}[\varphi(X_h(T))]| + |\mathbb{E}[\varphi(X_h(T)) - \mathbb{E}[\varphi(X_h^{\psi}(T))\Theta^{J_1}(T)]|.
\end{aligned}$$

The first part of the sum corresponds to the weak approximation error and for the second

summand, we obtain by Equation (5.3) that

$$|\mathbb{E}[\varphi(X_h(T))] - \mathbb{E}[\varphi(X_h^{\psi}(T))\Theta^{J_1}(T)]| = |\mathbb{E}[\varphi(X_h^{\psi}(T))(\Theta(T) - \Theta^{J_1}(T))]|.$$

By applying Cauchy-Schwarz inequality and using the assumption that the second moment of $\varphi(X_h^{\psi}(T))$ is bounded, we get

$$\begin{aligned}|\mathbb{E}[\varphi(X_h^{\psi}(T))(\Theta(T) - \Theta^{J_1}(T))]| &\leq \|\varphi(X_h^{\psi}(T))\|_{L^2(\Omega;\mathbb{R})}\|\Theta(T) - \Theta^{J_1}(T)\|_{L^2(\Omega;\mathbb{R})} \\ &\leq C\|\Theta(T) - \Theta^{J_1}(T)\|_{L^2(\Omega;\mathbb{R})}\end{aligned}$$

Due to the independence of the Brownian motions $\beta_k(t)$, Itô's isometry and the boundedness of $\kappa(s)$ we obtain

$$\begin{aligned}\|\Theta(T) - \Theta^{J_1}(T)\|^2_{L^2(\Omega;\mathbb{R})} &= \mathbb{E}\left[\left(\sum_{k=J_1+1}^{\infty}\int_0^T \sqrt{\mu_k}\kappa_k(s)\,\mathrm{d}\beta_k(s)\right)^2\right] \\ &= \sum_{k=J_1+1}^{\infty}\mu_k\,\mathbb{E}\left[\left(\int_0^T \kappa_k(s)\,\mathrm{d}\beta_k(s)\right)^2\right] \\ &= \sum_{k=J_1+1}^{\infty}\mu_k\,\mathbb{E}\left[\int_0^T \kappa_k(s)^2\,\mathrm{d}s\right] \\ &\leq C\sum_{k=J_1+1}^{\infty}\mu_k.\end{aligned}$$

Thus, we get

$$|\mathbb{E}[\varphi(X(T))] - \mathbb{E}[\varphi(X_h^{\psi}(T))\Theta^{J_1}(T)]|^2 \leq |\mathbb{E}[\varphi(X(T))] - \mathbb{E}[\varphi(X_h(T))]|^2 + C\sum_{k=J_1+1}^{\infty}\mu_k.$$

Finally, the weak convergence order $\mathcal{O}(h^{\alpha})$ is preserved if $\sum_{k=J_1+1}^{\infty}\mu_k = \mathcal{O}(h^{2\alpha})$. □

Until now we have discussed how to choose two different truncation parameters: First, the truncation parameter J_0 for the Karhunen-Loève expansion in Equation (5.5) that is chosen in a way such that certain convergence properties of the numerical methods are preserved, and second, the truncation parameter J_1 for the approximation of the density process, for which we found a specific selection criteria in terms of the weak convergence properties. Note that both parameters might be in general different depending on the numerical method, of which the properties have to be preserved. However, we subsequently choose $J = \max\{J_0, J_1\}$ for both truncations, which essentially means that our simulation results are based on J independent Brownian motions $\beta_k(t), k = 1, \ldots, J$.

Concerning the implementation of the density process of the two measure transformation examples from Section 3.1, there is a slight difference. The time-constant weight

operator defined in Equation (3.9) leads to an explicitly solvable SODE (5.7) with solution

$$\Theta^J(t) = \exp\left(\sum_{k=1}^{J} \sqrt{\mu_k}\psi_k\langle 1, e_k\rangle_H \beta_k(t) - \frac{t}{2}\sum_{k=1}^{J}\mu_k\psi_k^2\langle 1, e_k\rangle_H^2\right).$$

In contrast, the density process $\Theta(t)$ induced by the path-dependent weight operator given in Equation (3.10) is now represented as the solution of a linear one-dimensional SODE driven by J independent Brownian motion, i.e.

$$\mathrm{d}\Theta^J(t) = \Theta^J(t)\left(c\sum_{k=1}^{J}\sqrt{\mu_k}\langle \overline{G}(X^\psi(t))\eta, e_k\rangle_H \,\mathrm{d}\beta_k(t)\right), \quad \Theta^J(0) = 1, \tag{5.8}$$

which can be approximated by using a standard numerical scheme for SDEs such as e.g. the Euler-Maruyama scheme or the Milstein scheme.

6 Numerical experiments

In this section we present numerical experiments for the one-dimensional stochastic heat equation with homogeneous Dirichlet boundary conditions on the spatial domain $D = (0, 1)$ given by

$$\begin{aligned} \mathrm{d}X(t,x) &= \Delta_x X(t,x)\,\mathrm{d}t + G(X(t,x))\,\mathrm{d}W(t,x), \\ X(0,x) &= \sin(\pi x), && \text{for } x \in D, \\ X(t,0) &= X(t,1) = 0, && \text{for } t \in [0,T]. \end{aligned} \tag{6.1}$$

In the numerical experiments below, the Nemytskii operator $G : H \to L_{HS}(\mathcal{H}_0, H)$ is induced by either $\gamma(x) = \sin(x)$ (a globally bounded and Lipschitz continuous function) or by $\gamma(x) = \sigma x, \sigma \in \mathbb{R}$ (a globally unbounded, but Lipschitz continuous function). Note that the underlying Hilbert space is $H = L^2(D)$. Furthermore, we assume that the eigenvalues of the covariance operator Q of the Q-Wiener process $W(t), t \in [0, T]$, are defined by $\mu_k = k^{-(2r+1+\epsilon)}$ for given regularity parameter $r > 0$ and $\epsilon > 0$. Finally, we assume that Q and the operator A commute.

In this section we consider importance sampling techniques for SPDEs based on the path-dependent measure transformation from Equation (3.10) with $\eta = e_1 \in \mathcal{H}_0$. For the discretisation in space and time, we follow the methods proposed in Section 5.2, i.e. a standard finite element method using piecewise linear test functions as basis functions of V_h and for the time integration we use the backward Euler scheme on an equidistant partition θ_K of the time interval $[0, T]$ with time step size $\Delta t > 0$.

In Figure 2, we compare the performance of the importance sampling method with respect to different values of the scaling parameter $c \in \mathbb{R}$, where the diffusion operator is defined as the bounded Nemytskii operator G induced by $\gamma(x) = \sin(x)$. The functional $\varphi : H \to \mathbb{R}$ is chosen to be $\varphi(h) = \|h\|_H^2$ for all $h \in H$. In all experiments the specific choice $c = 0$ corresponds to the standard Monte Carlo estimator without using importance

sampling. The approximation of the solution is computed on a coarse spatial grid consisting of $N = 2^3$ grid points and on a partition θ_K of the time interval $[0, 1]$ consisting of $K = 2^6$ time steps. The regularity parameter controlling the decay of the eigenvalues of the covariance operator of the Q-Wiener process is chosen to be $r = 2$ and $\epsilon = 10^{-5}$.

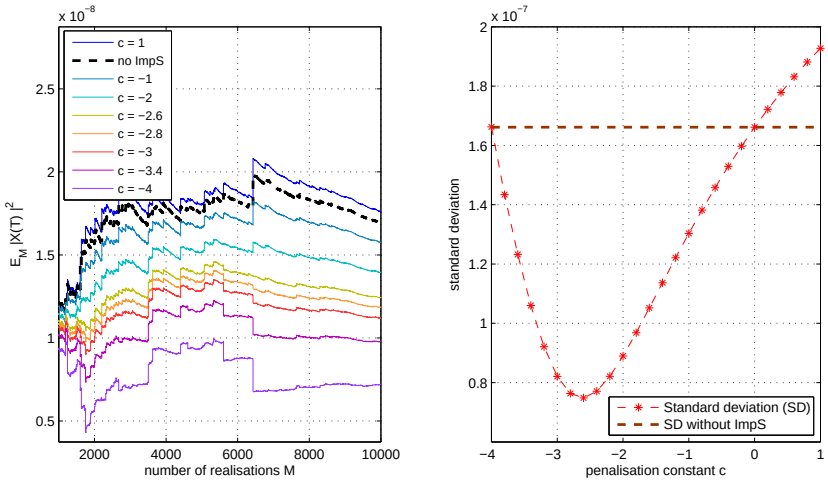


Figure 2: Simulation of $\mathbb{E}[\|X(1)\|_H^2]$ of the stochastic heat equation (6.1) with diffusion operator G defined as a Nemytskii operator based on $\gamma(x) = \sin(x)$.
Left: Evolution of Monte Carlo estimators over number of realisations (M) using importance sampling based on path-dependent weight operator from (3.10) with different scaling parameter c.
Right: Standard deviation of the simulated realisations ($M = 10^5$) over different values of scaling parameter c.

As we can see on the left hand side of Figure 2, there are jumps in the evolution of the standard Monte Carlo estimator (dashed line) due to realisations that are remarkably larger than the average. This causes severe difficulties for estimating $\mathbb{E}[\|X(1)\|_H^2]$ since we are likely to overestimate the mean-square process right after such jumps. For scaling parameters $c < 0$ we damp these jumps and for this reason we are able to reduce the variance of the simulated realisations. However if we choose c too small, then *negative* jumps in the evolution of the Monte Carlo estimator appear. These jumps are caused by numerical instabilities in the simulation of $\Theta(t)$, which result in negative realisations of the simulated density process. As a consequence, one has to choose a scaling parameter c that leads to variance reduction, but still guarantees numerical stability of the involved approximation schemes.

In Figure 3, we consider a realisation of the density process $\Theta(T)$ based on the path-dependent weight operator given in Equation (3.10) for fixed scaling parameter $c = -2$. As proposed in Section 5 we truncate the infinite series appropriately (due to the regularity parameter $r = 2$ it is sufficient to choose the truncation parameter $J = N$ according to Lemma 5.4), which leads to the one-dimensional SDE (5.8) that is approximated by using the standard Euler-Maruyama scheme.

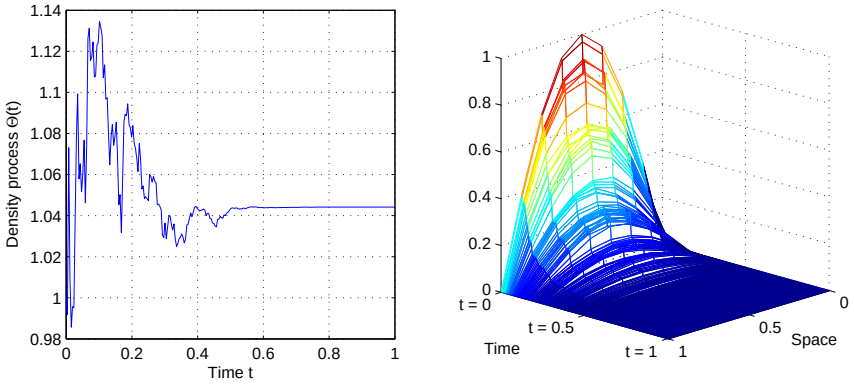


Figure 3: *Left*: A realisation of the approximated density process $\Theta^J(t), t \in [0,1]$, given in Equation (5.8) approximated by the standard Euler-Maruyama scheme. *Right*: The underlying realisation of the approximation of the transformed solution $X^{\psi}(t), t \in [0,1]$.

All the simulations in Figure 3 are performed by using $K = 2^8$ time steps in order to observe the path-dependence of $\Theta(t), t \in [0,1]$. After a short time horizon (here $t \approx 0.5$) the density process stays almost constant due to the fact that the values of the underlying trajectory of $X^{\psi}(t)$ are very small. This implies that the realisation of the transformed process at the end time point is weighted by a constant that is mainly dependent on time intervals, where the norm of the transformed process is large or at least significantly larger than 0. As a consequence of the chosen scaling by the parameter c, we are again forcing ($c > 0$) or damping ($c < 0$) the importance of such events for the Monte Carlo estimator.

In the following part, we consider simulations of SPDE (2.1), where the diffusion operator G is defined as a Nemytskii operator based on $\gamma(x) = \sigma x, \sigma \in \mathbb{R}$. Thus, as proposed in Section 3.1 for the path-dependent measure transformation, we have to derive a representation of the truncated operator $\overline{G}$ in order to guarantee the martingale property of the density process $\Theta(t)$. According to the results from Section 3.1 this can be achieved for Nemytskii operators by approximating the underlying function $\gamma(x)$ by a bounded function $\overline{\gamma} : \mathbb{R} \to [-C_G, C_G]$. Note that there are different possibilities how to approximate $\overline{\gamma}$. First, one can truncate γ if its absolute value exceeds a certain threshold $C_G > 0$, i.e. choose $\overline{\gamma}$ as

$$\overline{\gamma}(x) = \gamma_c(C_G; x) = \operatorname{sign}(\gamma(x)) \min\{C_G, |\gamma(x)|\} \qquad \text{for all } x \in \mathbb{R}. \tag{6.2}$$

The choice of C_G depends on typical solution properties, e.g., one could choose C_G such that the majority of the discretised solution is smaller in absolute value than the chosen constant C_G. In Figure 4, we consider the case $\gamma(x) = x$ and the bounded function γ_c is truncated at $C_G = 1$. Further details concerning the numerical outcome of the Monte Carlo simulations using the truncated function γ_c can be found in Table 1.

According to Figure 4, we are also able to reduce the variance in the case of a globally unbounded diffusion operator. However, for a large and negative scaling parameter c, i.e. for c being approximately smaller than -2.4, the performance of the importance sam-

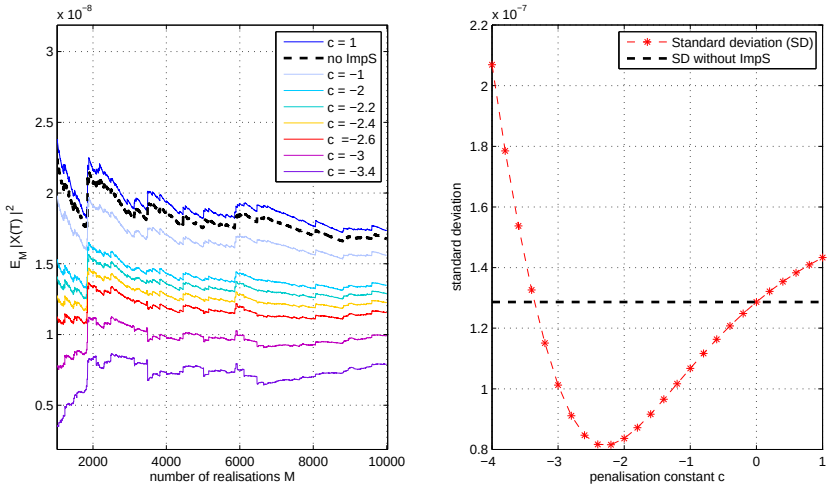


Figure 4: Simulation of $\mathbb{E}[\|X(1)\|_H^2]$ of the stochastic heat equation (6.1) with diffusion operator G defined as a Nemytskii operator based on $\gamma(x) = x$.
Left: Evolution of Monte Carlo estimators over M using importance sampling based on path-dependent weight operator from (3.10) with different scaling parameter c.
Right: Standard deviation of the simulated realisations ($M = 10^5$) over different values of scaling parameter c.

pling method is worse compared to the case of the bounded diffusion ($\gamma(x) = \sin(x)$). This happens due to the fact that the same numerical instabilities in the simulation of the density process as in Figure 2 already appear for choices $c \approx -2.4$.

By the same arguments as they were used to derive γ_c in Equation (6.2), one could choose $\overline{\gamma}$ as a bounded function such that the function γ is approximated well by $\overline{\gamma}$ at typical values of the (discretised) quantity of interest. For instance for $\gamma(x) = x$ one could consider an approximation by a sigmoid function, e.g. $\overline{\gamma}(x) = \mathrm{erf}(x)$.

In Table 1, we compare numerical tests based on different importance sampling techniques using the path-dependent density operator from Section 3.1. Each of the 4 tests is based on $M = 10^5$ independent realisations. As a measure of the effectiveness of the importance sampling methods we consider the ratio between the standard deviation of the transformed processes (SD_c) over the standard deviation (SD_0) of the original realisations of $\|X(1)\|_H^2$. This ratio measures the relative change of the Monte Carlo error using importance sampling based on scaling parameter c, i.e. for values $SD_c/SD_0 < 1$ the importance sampling method performs better than the original Monte Carlo simulation and for values c, for which $SD_c/SD_0 > 1$, the performance is worse.

Hence, the proposed importance sampling techniques can reduce the Monte Carlo error significantly for different scaling parameters c. For the specific value $c = -2$ we achieve a remarkable reduction of the variance in all numerical experiments. Thus, these results show the efficiency of the proposed methods, since in order to obtain the same accuracy as for the standard Monte Carlo estimator, one needs $M_c = (\frac{SD_c}{SD_0})^2 M$ samples.

Table 1: Performance of the importance sampling techniques applied to SPDE (2.1) with path-dependent measure transformation from Section 3.1 using different scaling parameter c for test function $\varphi(h) = \|h\|_H^2, h \in H$.

	Test 1: $\gamma(x) = \sin(x)$		Test 2: $\gamma(x) = x$ with $\overline{\gamma}(x) = \gamma_c(1; x)$	
c	SD_c	SD_c/SD_0	SD_c	SD_c/SD_0
1	1.928E-07	1.1607	1.433E-07	1.1143
0	1.661E-07	1	1.286E-07	1
-1	1.303E-07	0.7845	1.067E-07	0.8297
-2	8.897E-08	0.5356	8.362E-08	0.6502
-3	8.210E-08	0.4943	9.111E-08	0.7085
-4	1.662E-07	1.0006	2.069E-07	1.6089
	Test 3: $\gamma(x) = x$ with $\overline{\gamma}(x) = \gamma_c(0.5; x)$		Test 4: $\gamma(x) = x$ with $\overline{\gamma}(x) = \mathrm{erf}(x)$	
c	SD_c	SD_c/SD_0	SD_c	SD_c/SD_0
1	2.467E-07	1.1989	2.516E-07	1.1770
0	2.060E-07	1	2.138E-07	1
-1	1.510E-07	0.7332	1.581E-07	0.7395
-2	1.209E-07	0.5870	1.242E-07	0.5810
-3	2.261E-07	1.0977	2.555E-07	1.1948
-4	4.630E-07	2.2481	5.690E-07	2.6611

References

[1] M. Ableidinger, E. Buckwar, and A. Thalhammer. “An importance sampling technique in Monte Carlo methods for SDEs with a.s. stable and mean-square unstable equilibrium”. In: *J. Comput. Appl. Math.* 316 (2017), pp. 3–14.

[2] A. Barth and A. Lang. “Multilevel Monte Carlo method with applications to stochastic partial differential equations”. In: *Int. J. Comput. Math.* 89.18 (2012), pp. 2479–2498.

[3] A. Barth and A. Lang. “Simulation of stochastic partial differential equations using finite element methods”. In: *Stochastics* 84.2-3 (2012), pp. 217–231.

[4] A. Barth, A. Lang, and C. Schwab. “Multilevel Monte Carlo method for parabolic stochastic partial differential equations”. In: *BIT* 53.1 (2013), pp. 3–27.

[5] Z. Brzeźniak et al. “Itô’s formula in UMD Banach spaces and regularity of solutions of the Zakai equation”. In: *J. Differential Equations* 245.1 (2008), pp. 30–58.

[6] G. Da Prato, M. Röckner, and A. Jentzen. “A mild Itô formula for SPDEs”. eprint: arXiv:1009.3526 (2011), 31 pages.

[7] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and its Applications, 44. Cambridge University Press, Cambridge, 1992.

[8] L. Gawarecki and V. Mandrekar. *Stochastic differential equations in infinite dimensions with applications to stochastic partial differential equations*. Probability and its Applications (New York). Springer, Heidelberg, 2011.

[9] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Vol. 53. Springer Science & Business Media, 2003.

[10] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Applications of Mathematics, 23. Springer-Verlag, Berlin, 1992.

[11] R. Kruse. *Strong and Weak Approximation of Semilinear Stochastic Evolution Equations*. Lecture Notes in Mathematics, 2093. Springer, Cham, 2014.

[12] K. Liu. *Stability of Infinite Dimensional Stochastic Differential Equations with Applications*. Chapman & Hall/CRC Monographs, Surveys in Pure, and Applied Mathematics 135, 2006.

[13] W. Liu and M. Röckner. *Stochastic partial differential equations: an introduction*. Universitext. Springer, Cham, 2015.

[14] G. J. Lord, C. E. Powell, and T. Shardlow. *An Introduction to Computational Stochastic PDEs*. Cambridge Texts in Applied Mathematics. Cambridge University Press, New York, 2014.

[15] G. N. Milstein and M. V. Tretyakov. “Practical variance reduction via regression for simulating diffusions”. In: *SIAM Journal on Numerical Analysis* 47.2 (2009), pp. 887–910.

[16] G. N. Milstein and M. V. Tretyakov. “Solving parabolic stochastic partial differential equations via averaging over characteristics”. In: *Math. Comp.* 78.268 (2009), pp. 2075–2106.

[17] G. N. Milstein and M. V. Tretyakov. *Stochastic Numerics for Mathematical Physics*. Scientific Computation. Springer-Verlag, Berlin, 2004.

[18] B. Øksendal. *Stochastic differential equations. An introduction with applications*. Fourth Edition. Universitext. Springer-Verlag, Berlin, 1995.

[19] C. Prévôt and M. Röckner. *A concise course on stochastic partial differential equations*. Vol. 1905. Lecture Notes in Mathematics. Springer, Berlin, 2007.

[20] P. E. Protter. *Stochastic integration and differential equations*. Second Edition. Vol. 21. Stochastic Modelling and Applied Probability. Springer-Verlag, Berlin, 2005.

4 Mean-square stability analysis of stochastic differential equations in infinite dimensions

Annika Lang, Andreas Petersson, and Andreas Thalhammer

The (asymptotic) behaviour of the second moment of solutions to stochastic differential equations is treated in mean-square stability analysis. The purpose of this article is to discuss this property for approximations of infinite-dimensional stochastic differential equations and give necessary and sufficient conditions that ensure mean-square stability of the considered finite-dimensional approximations. Stability properties of typical discretization schemes such as combinations of spectral Galerkin, finite element, Euler–Maruyama, Milstein, Crank–Nicolson, and forward and backward Euler methods are characterized. Furthermore, results on their relationship to stability properties of the analytical solutions are provided. Simulations of the stochastic heat equation illustrate the theory.

Key words. Asymptotic mean-square stability, numerical approximations of stochastic differential equations, linear stochastic partial differential equations, Lévy processes, rational approximations, Galerkin methods, spectral methods, finite element methods, Euler–Maruyama scheme, Milstein scheme.

AMS subject classifications. 60H15, 65M12, 60H35, 65C30, 65M60

1 Introduction

In many fields of science, stochastic differential equations (SDEs) in infinite dimensions often in form of stochastic partial differential equations (SPDEs) are used to model problems with uncertainties. Since many of the considered equations can only be solved numerically, the numerical analysis of such SDEs has gained a lot of attention in the literature over the last two decades. Most of the numerical methods used for approximating the infinite-dimensional solution process $X(t)$ are based on a discretization in space by means of a family $(V_h, h \in (0,1])$ of finite-dimensional subspaces of the original solution space (with refinement parameter h) and on a time discretization with step size Δt. For many of these numerical methods, strong and partly also weak convergence results are available in the literature, (see, e.g., [19] for an overview), where the authors investigate the asymptotic behaviour with respect to $h, \Delta t \to 0$. However, for numerical simulations we are restricted to a fixed refinement parameter h in space and a time step size $\Delta t > 0$. For this reason, we investigate in this article the asymptotic qualitative behaviour of approximations of $X(t)$

as time $t \to \infty$ for fixed refinement parameters $h > 0$ and $\Delta t > 0$.

We focus on the asymptotic mean-square stability analysis of approximations of linear SDEs attaining values in a real separable Hilbert space H, i.e., we consider the asymptotic behaviour of $\mathbb{E}[\|X_h^j\|_H^2]$ as $j \to \infty$, where X_h^j denotes an approximation of $X(t_j) = X(j\Delta t)$. A lot of effort has been dedicated to the asymptotic mean-square stability analysis of solutions of SDEs in finite dimensions, see e.g., [16, 25, 2], as well as in infinite dimensions, see e.g., [23]. The mean-square stability analysis of numerical approximations of SDEs started by considering the approximations of the one-dimensional geometric Brownian motion, see e.g., [28, 13, 14]. As it has been pointed out in [9, 10], the mean-square stability analysis of linear SDE systems in higher dimensions and their approximations is also necessary, since the asymptotic behaviour of the corresponding mean-square processes of systems with commuting and non-commuting matrices are in general not equal.

The tools to perform mean-square stability analysis of SDE approximations presented in [10] can also be used for approximations of infinite-dimensional SDEs by a method of lines approach: After a spatial discretization (by, e.g., a Galerkin finite element method) the mean-square stability properties of the resulting finite-dimensional system of stochastic differential equations and the corresponding approximations can be determined by considering the eigenvalues of $(N_h^2 \times N_h^2)$-dimensional matrices, where N_h denotes the dimension of V_h. However, due to the computational complexity, neither the symbolic nor the numerical computation of these eigenvalues can be done for arbitrarily large systems.

For this reason, we extend the results from [10] to linear operators based on approximation schemes applied to H-valued SDEs driven by square-integrable, càdlàg martingales by using representations of the tensor-product-space-valued process $\mathbb{E}[X_h^j \otimes X_h^j]$. By applying these results to space approximations with Galerkin methods in Section 4, we show sufficient conditions for the asymptotic mean-square stability of different time discretization schemes combined with Euler–Maruyama and Milstein methods.

From an application point of view, mean-square stability analysis of approximations of H-valued SDEs is very important, since there exist frequently used numerical techniques for approximating paths or moments of infinite-dimensional solution trajectories that require specific mean-square stability properties of the numerical solution. Here, we mention the multilevel Monte Carlo (MLMC) estimator which is known to be a computationally efficient method to estimate quantities of interest $\mathbb{E}[\varphi(X(T))]$, see, e.g., [5, 3] for details. This estimator uses realizations of the solution process on a hierarchical collection of spatial and temporal refinement levels. Consequently, it is necessary to use a sufficiently (mean-square) stable numerical approximation scheme on all levels (especially on the coarsest levels) to get a reasonable approximation of $\mathbb{E}[\varphi(X(T))]$, see, e.g., [1]. We provide conditions under which the infinite-dimensional solution $X(t)$ and its approximations based on backward Euler schemes share the same mean-square stability properties without any restriction on the refinement parameters h and Δt.

The outline of this article is as follows: In Section 2 we set up the framework in which mean-square stability is considered. For this, we recall basic results on linear H-valued

SDEs driven by square-integrable, càdlàg martingales and their approximations. Section 3 contains the theory of mean-square stability analysis for approximations of infinite-dimensional SDEs in an abstract way. In the main result, necessary and sufficient conditions for asymptotic mean-square stability are derived. These results are then applied in Section 4 to numerical approximations based on spatial Galerkin discretization schemes and time discretizations based on backward/forward Euler and Crank–Nicolson schemes combined with Euler–Maruyama and Milstein methods. We conclude this work by showing numerical experiments in Section 5 that illustrate the theory.

2 Hilbert space-valued SDE and approximations

In this section we set up the framework for mean-square stability. The theoretical setting is based on [26] to which the reader is referred for further details. Throughout, let U and H be real separable Hilbert spaces with inner products denoted by $\langle\cdot,\cdot\rangle_U$ and $\langle\cdot,\cdot\rangle_H$, respectively, and let $(\Omega,\mathcal{A},(\mathcal{F}_t,t\geq 0),P)$ be a complete filtered probability space satisfying the "usual conditions". We consider in what follows the linear stochastic differential equation on $\mathbb{R}_+$

$$\mathrm{d}X(t)=(AX(t)+F(X(t)))\,\mathrm{d}t+G(X(t))\,\mathrm{d}M(t) \tag{2.1}$$

with $\mathcal{F}_0$-measurable initial condition $X(0)=X_0\in L^2(\Omega;H)$, i.e., X_0 is an H-valued, square-integrable random variable. Here, $A:\mathcal{D}(A)\to H$ is the generator of a C_0-semigroup $S=(S(t),t\geq 0)$ on H and F is a linear and bounded operator on H, i.e., $F\in L(H)$. Let us further assume that $M=(M(t),t\geq 0)$ is a U-valued, càdlàg, square-integrable martingale which is adapted to the filtration and has independent increments, i.e., the filtration is *admissible* with respect to M. An example of such a filtration is the natural filtration $\mathcal{F}^M$ generated by M. Let us assume further that there exists a self-adjoint, positive semidefinite trace class operator $Q\in L(U)$ such that for all $t\geq s\geq 0$

$$\langle\langle M,M\rangle\rangle_t-\langle\langle M,M\rangle\rangle_s\leq(t-s)Q,$$

where the operator angle bracket process $(\langle\langle M,M\rangle\rangle_t,t\geq 0)$ is defined as

$$\langle\langle M,M\rangle\rangle_t=\int_0^t Q_s\,\mathrm{d}\,\langle M,M\rangle_s\,.$$

By $(\langle M,M\rangle_t,t\geq 0)$ we denote the angle bracket process from the Doob–Meyer decomposition and $(Q_t,t\geq 0)$ is the martingale covariance. Finally, let G be a linear mapping from H to $L_{\mathsf{HS}}(Q^{1/2}(U);H)$, the space of Hilbert–Schmidt mappings from $Q^{1/2}(U)$ to H. Then (2.1) admits by results in [26, Chapter 9] an up to modification unique mild solution which is càdlàg and for $t\geq 0$ given by

$$X(t)=S(t)X_0+\int_0^t S(t-s)F(X(s))\,\mathrm{d}s+\int_0^t S(t-s)G(X(s))\,\mathrm{d}M(s). \tag{2.2}$$

As a standard example in this context that is used throughout the manuscript, we introduce the stochastic heat equation.

Example 2.1. *Let the separable Hilbert space $H = L^2([0,1])$ be the space of square-integrable functions on $[0,1]$. On this space we consider the operator $A = \nu\Delta$, where $\nu > 0$ and Δ denotes the Laplace operator with homogeneous zero Dirichlet boundary conditions which is the generator of a C_0-semigroup, cf. [19, Example 2.21]. Furthermore, let the square-integrable martingale $M = L$ be a U-valued Lévy process. The equation*

$$\mathrm{d}X(t) = \nu\Delta X(t)\,\mathrm{d}t + G(X(t))\,\mathrm{d}L(t)$$

is referred to as the (homogeneous) stochastic heat equation.

The goal of this manuscript is to characterize mean-square stability properties of fully discrete approximations of solutions of (2.1) such as the mild solution (2.2), which we introduce next. Therefore, let $(V_h, h \in (0,1])$ be a family of finite-dimensional subspaces $V_h \subset H$ with $\dim(V_h) = N_h \in \mathbb{N}$ indexed by a refinement parameter h for the space approximation. With an inner product induced by $\langle\cdot,\cdot\rangle_H$, V_h becomes a Hilbert space with norm $\|\cdot\|_H$. For a linear operator $D : V_h \to V_h$, the operator norm $\|D\|_{L(V_h)}$ is therefore given by

$$\|D\|_{L(V_h)} = \sup_{v\in V_h} \frac{\|Dv\|_H}{\|v\|_H}$$

and can be seen to coincide with $\|DP_h\|_{L(H)}$, where P_h is the orthogonal projection onto V_h.

We choose time steps $t_j = j\Delta t$, $j \in \mathbb{N}_0$, for a fixed step size $\Delta t > 0$. Hence, $t \to \infty$ is equivalent to $j \to \infty$. For a given time point t_j, we denote an approximation of $X(t_j)$ in V_h by X_h^j, which is obtained by a numerical approximation scheme given by

$$X_h^{j+1} = D_{\Delta t,h}^{\mathrm{det}} X_h^j + D_{\Delta t,h}^{\mathrm{stoch},j} X_h^j \tag{2.3}$$

with initial condition X_h^0 approximating X_0. By $D_{\Delta t,h}^{\mathrm{det}} \in L(V_h)$ we denote a linear operator approximating the solution operator of the deterministic part of (2.1)

$$\mathrm{d}X(t) = (AX(t) + F(X(t)))\,\mathrm{d}t, \quad t \in [t_j, t_{j+1})$$

and by $D_{\Delta t,h}^{\mathrm{stoch},j}$ we denote an $L(V_h)$-valued random variable approximating the solution operator of the stochastic part

$$\mathrm{d}X(t) = G(X(t))\,\mathrm{d}M(t), \quad t \in [t_j, t_{j+1}).$$

We assume that the family of stochastic approximation operators $(D_{\Delta t,h}^{\mathrm{stoch},j}, j \in \mathbb{N}_0)$ is $\mathcal{F}$-compatible in the sense of [11, 20], i.e., for given $h, \Delta t > 0$, $D_{\Delta t,h}^{\mathrm{stoch},j}$ is $\mathcal{F}_{t_{j+1}}$-measurable and $\mathbb{E}[D_{\Delta t,h}^{\mathrm{stoch},j}|\mathcal{F}_{t_j}] = 0$ for all $j \in \mathbb{N}_0$, where we denote by $\mathbb{E}[\cdot|\mathcal{G}]$ the conditional expectation with respect to a σ-algebra $\mathcal{G}$. Throughout the article, it is also assumed that for any given

$h, \Delta t > 0, \|D_{\Delta t,h}^{\mathrm{stoch},j}\|_{L^2(\Omega;L(V_h))} = \mathbb{E}[\|D_{\Delta t,h}^{\mathrm{stoch},j}\|_{L(V_h)}^2]^{1/2} < \infty$ for all $j \in \mathbb{N}_0$.

Additionally, we assume that for all $j \in \mathbb{N}_0$, the operator $D_{\Delta t,h}^{\mathrm{stoch},j}$ approximating the stochastic integral on the time interval $[t_j, t_{j+1})$ is independent of $\mathcal{F}_{t_j}$. This assumption is motivated by the fact that the increments of the process M denoted by $\Delta M^j = M(t_{j+1}) - M(t_j)$, are assumed to be independent of $\mathcal{F}_{t_j}$ for all $j \in \mathbb{N}_0$. Nevertheless, we remark that it would be sufficient to assume the rather technical condition

$$\mathbb{E}\left[(D_{\Delta t,h}^{\mathrm{stoch},j} \otimes D_{\Delta t,h}^{\mathrm{stoch},j})(X_h^j \otimes X_h^j)\right] = \mathbb{E}\left[D_{\Delta t,h}^{\mathrm{stoch},j} \otimes D_{\Delta t,h}^{\mathrm{stoch},j}\right] \mathbb{E}[X_h^j \otimes X_h^j].$$

3 Asymptotic mean-square stability analysis

Let us recall the main definitions of (asymptotic) mean-square stability theory for SDE (2.1) before we derive tools for the mean-square stability analysis of the corresponding numerical approximations (2.3).

For this, we examine the qualitative behaviour of an equilibrium (solution), which is defined as a constant solution $X_{\mathrm{e}}(t) = x_{\mathrm{e}} \in H$ for all $t \geq 0$ of (2.1) satisfying $(A+F)x_{\mathrm{e}} = 0$ and $G(x_{\mathrm{e}}) = 0$. Note that for linear SPDEs such as the one given in (2.1), the zero initial value $X_0 = 0$ admits the *zero* or *trivial solution* $X_{\mathrm{e}}(t) = 0$ for all $t > 0$, and therefore, the zero solution is an equilibrium solution of (2.1). As it is shown, e.g., in [23], it is sufficient to consider only the stability properties of the zero solution, since – also in the non-linear case – the zero solution is an equilibrium of the SDE corresponding to $(Y(t) = X(t) - X_{\mathrm{e}}(t), t \geq 0)$ for any equilibrium $X_{\mathrm{e}}(t)$ of $X(t)$.

Due to these considerations, we are interested in the qualitative behaviour of the zero solution of the underlying linear SDE. In the following definition, the main stability concept used in this work, i.e., the (asymptotic) mean-square stability of the zero solution, is introduced:

Definition 3.1. Let $X(t)$ be a solution of (2.1). Then the zero solution $(X_{\mathrm{e}}(t) = 0, t \geq 0)$ of (2.1) is called *mean-square stable* if, for every $\varepsilon > 0$, there exists $\delta > 0$ such that for all $t \geq 0$,

$$\mathbb{E}[\|X(t)\|_H^2] < \varepsilon$$

whenever $\mathbb{E}[\|X_0\|_H^2] < \delta$.

It is called *asymptotically mean-square stable* if it is mean-square stable and there exists $\delta > 0$ such that $\mathbb{E}[\|X_0\|_H^2] < \delta$ implies

$$\lim_{t\to\infty} \mathbb{E}[\|X(t)\|_H^2] = 0.$$

Furthermore, it is called *asymptotically mean-square unstable* if it is not asymptotically mean-square stable.

Similarly, an equilibrium (solution) of the approximation scheme (2.3) is a constant solution $X_{h,\mathrm{e}}^j = x_{h,\mathrm{e}} \in V_h$ for all $j \in \mathbb{N}_0$ for which $D_{\Delta t,h}^{\mathrm{det}} x_{h,\mathrm{e}} = 0$ and $D_{\Delta t,h}^{\mathrm{stoch},j} x_{h,\mathrm{e}} = 0$ for

all $j \in \mathbb{N}_0$ holds. By the same arguments as for (2.1), we are interested in the qualitative behaviour of the zero solution of (2.3), where the zero solution is the solution $X^j_{h,\mathrm{e}} = 0$ of (2.3) that is induced by the zero initial value $X^0_h = 0$. We define mean-square stability of the zero solution of (2.3) in what follows, which can be seen as the discrete version of Definition 3.1.

Definition 3.2. Let $X_h = (X^j_h, j \in \mathbb{N}_0)$ be the numerical approximation given by (2.3) for fixed time step size Δt and refinement parameter h. Then the zero solution $(X^j_{h,\mathrm{e}} = 0, j \in \mathbb{N}_0)$ of (2.3) is called *mean-square stable* if, for every $\varepsilon > 0$, there exists $\delta > 0$ such that, for all $j \in \mathbb{N}_0$,

$$\mathbb{E}[\|X^j_h\|^2_H] < \varepsilon$$

whenever $\mathbb{E}[\|X^0_h\|^2_H] < \delta$.

It is called *asymptotically mean-square stable* if it is mean-square stable and there exists $\delta > 0$ such that $\mathbb{E}[\|X^0_h\|^2_H] < \delta$ implies

$$\lim_{j\to\infty} \mathbb{E}[\|X^j_h\|^2_H] = 0.$$

Furthermore, it is called *asymptotically mean-square unstable* if it is not asymptotically mean-square stable.

When applied to $Y_j = X^j_h$, the following lemma provides an equivalent condition for the mean-square stability of the zero solution of the fully discrete scheme (2.3) in terms of the tensor-product-space-valued process $X^j_h \otimes X^j_h \in V^{(2)}_h$, where $\otimes$ denotes the Hilbert tensor product and for a general Hilbert space H, the abbreviation $H^{(2)} = H \otimes H$ is used. More precisely, $H \otimes H$ is defined as the completion of the algebraic tensor product with respect to the norm induced by the inner product

$$\langle v, w\rangle_{H\otimes H} = \sum_{i=1}^{N}\sum_{j=1}^{M} \langle v_{1,i}, w_{1,j}\rangle_H \langle v_{2,i}, w_{2,j}\rangle_H$$

where $v = \sum_{i=1}^N v_{1,i} \otimes v_{2,i}$ and $w = \sum_{j=1}^M w_{1,j} \otimes w_{2,j}$ are representations of elements v, w in the algebraic tensor product.

Lemma 3.3. *Let V_h be any finite-dimensional subspace of H. Then, for any sequence $(Y_j, j \in \mathbb{N}_0)$ of V_h-valued, square-integrable random variables, $\lim_{j\to\infty} \mathbb{E}[Y_j \otimes Y_j] = 0$ if and only if $\lim_{j\to\infty} \mathbb{E}[\|Y_j\|^2_H] = 0$.*

Proof. It is clear that $\lim_{j\to\infty} \mathbb{E}[Y_j \otimes Y_j] = 0$ is equivalent to $\lim_{j\to\infty} \|\,\mathbb{E}[Y_j \otimes Y_j]\|_{H^{(2)}} = 0$, since H and $H^{(2)}$ are Hilbert spaces with norm induced by the inner product.

By Parseval's identity, for an orthonormal basis $(\psi_1, \ldots, \psi_{N_h})$ of V_h we have

$$\left\|\mathbb{E}\left[Y_j \otimes Y_j\right]\right\|^2_{H^{(2)}} = \sum_{k,\ell=1}^{N_h} \left|\mathbb{E}\left[\langle Y_j \otimes Y_j, \psi_k \otimes \psi_\ell\rangle_{H^{(2)}}\right]\right|^2 = \sum_{k,\ell=1}^{N_h} \left|\mathbb{E}\left[\langle Y_j, \psi_k\rangle_H \langle Y_j, \psi_\ell\rangle_H\right]\right|^2$$

and similarly

$$\mathbb{E}\left[\|Y_j\|_H^2\right] = \sum_{k=1}^{N_h} \mathbb{E}\left[\langle Y_j, \psi_k\rangle_H^2\right].$$

Therefore, we immediately obtain one implication, while the other follows directly from the fact that

$$\|\mathbb{E}\left[Y_j \otimes Y_j\right]\|_{H^{(2)}} \leq \mathbb{E}\left[\|Y_j \otimes Y_j\|_{H^{(2)}}\right] = \mathbb{E}\left[\|Y_j\|_H^2\right]. \qquad \square$$

This lemma enables us to show the following sufficient condition for the asymptotic mean-square stability of the zero solution of the discrete approximation scheme (2.3).

Theorem 3.4. *Let* $X_h = (X_h^j, j \in \mathbb{N}_0)$ *be given by* (2.3)*. Furthermore, let*

$$\mathcal{S}_j = D_{\Delta t,h}^{\det} \otimes D_{\Delta t,h}^{\det} + \mathbb{E}[D_{\Delta t,h}^{\mathrm{stoch},j} \otimes D_{\Delta t,h}^{\mathrm{stoch},j}].$$

Then the zero solution of (2.3) *is asymptotically mean-square stable, if*

$$\lim_{j\to\infty} \|\mathcal{S}_j \cdots \mathcal{S}_0\|_{L(V_h^{(2)})} = 0.$$

Proof. Let us first remark that it is clear that $\mathcal{S}_j \in L(V_h^{(2)})$ for all $j \in \mathbb{N}_0$ by the properties of $D_{\Delta t,h}^{\det}$ and $D_{\Delta t,h}^{\mathrm{stoch},j}$ and the properties of the Hilbert tensor product. In order to show asymptotic mean-square stability of the zero solution of (2.3), it suffices to show

$$\mathbb{E}[X_h^j \otimes X_h^j] \to 0$$

as $j \to \infty$ by Lemma 3.3. For this, consider

$$\begin{aligned}\mathbb{E}[X_h^{j+1} \otimes X_h^{j+1}] &= \mathbb{E}\left[(D_{\Delta t,h}^{\det} + D_{\Delta t,h}^{\mathrm{stoch},j})X_h^j \otimes (D_{\Delta t,h}^{\det} + D_{\Delta t,h}^{\mathrm{stoch},j})X_h^j\right] \\ &= \mathbb{E}\left[(D_{\Delta t,h}^{\det} \otimes D_{\Delta t,h}^{\det})(X_h^j \otimes X_h^j)\right] + \mathbb{E}\left[(D_{\Delta t,h}^{\mathrm{stoch},j} \otimes D_{\Delta t,h}^{\mathrm{stoch},j})(X_h^j \otimes X_h^j)\right] \\ &\quad + \mathbb{E}\left[(D_{\Delta t,h}^{\det} \otimes D_{\Delta t,h}^{\mathrm{stoch},j})(X_h^j \otimes X_h^j)\right] + \mathbb{E}\left[(D_{\Delta t,h}^{\mathrm{stoch},j} \otimes D_{\Delta t,h}^{\det})(X_h^j \otimes X_h^j)\right].\end{aligned}$$

Since $D_{\Delta t,h}^{\mathrm{stoch},j}$ is independent of $\mathcal{F}_{t_j}$, the mixed terms vanish by the observation that

$$\begin{aligned}\mathbb{E}\left[(D_{\Delta t,h}^{\det} \otimes D_{\Delta t,h}^{\mathrm{stoch},j})(X_h^j \otimes X_h^j)\right] &= \mathbb{E}\left[D_{\Delta t,h}^{\det} \otimes D_{\Delta t,h}^{\mathrm{stoch},j}\right] \mathbb{E}[X_h^j \otimes X_h^j] \\ &= \left(D_{\Delta t,h}^{\det} \otimes \mathbb{E}[D_{\Delta t,h}^{\mathrm{stoch},j}]\right) \mathbb{E}[X_h^j \otimes X_h^j] = 0,\end{aligned}$$

where we used that $D_{\Delta t,h}^{\mathrm{stoch},j}$ is $\mathcal{F}$-compatible which implies that

$$\mathbb{E}[D_{\Delta t,h}^{\mathrm{stoch},j}] = \mathbb{E}\left[\mathbb{E}[D_{\Delta t,h}^{\mathrm{stoch},j}|\mathcal{F}_{t_j}]\right] = 0.$$

Hence,

$$\mathbb{E}[X_h^{j+1} \otimes X_h^{j+1}] = \mathbb{E}\left[\left(D_{\Delta t,h}^{\det} \otimes D_{\Delta t,h}^{\det} + D_{\Delta t,h}^{\mathrm{stoch},j} \otimes D_{\Delta t,h}^{\mathrm{stoch},j}\right)(X_h^j \otimes X_h^j)\right].$$

By once again noting that $D_{\Delta t,h}^{\mathrm{stoch},j}$ is independent of $\mathcal{F}_{t_j}$ and that the linear operator $D_{\Delta t,h}^{\det}$ is deterministic, we get

$$\mathbb{E}[X_h^{j+1} \otimes X_h^{j+1}] = \mathcal{S}_j\, \mathbb{E}[X_h^j \otimes X_h^j] = (\mathcal{S}_j \cdots \mathcal{S}_0)\, \mathbb{E}[X_h^0 \otimes X_h^0].$$

Since $\lim_{j\to\infty} \|\mathcal{S}_j \cdots \mathcal{S}_0\|_{L(V_h^{(2)})} = 0$ implies that there exists a constant C such that

$$\sup_{j\in\mathbb{N}} \|\mathcal{S}_j \cdots \mathcal{S}_0\|_{L(V_h^{(2)})} \le C,$$

it holds that

$$\begin{aligned}
\mathbb{E}[\|X_h^{j+1}\|_H^2]^2 &= \left(\sum_{k=1}^{N_h} \mathbb{E}[\langle X_h^{j+1}, \psi_k\rangle_H^2]\right)^2 \le N_h \left(\sum_{k=1}^{N_h} \mathbb{E}[\langle X_h^{j+1}, \psi_k\rangle_H^2]^2\right) \\
&\le N_h \big\|\mathbb{E}[X_h^{j+1} \otimes X_h^{j+1}]\big\|_{H^{(2)}}^2 \le N_h \left(\|\mathcal{S}_j \cdots \mathcal{S}_0\|_{L(V_h^{(2)})} \mathbb{E}[\|X_h^0\|_H^2]\right)^2 \\
&\le C^2 N_h\, \mathbb{E}[\|X_h^0\|_H^2]^2.
\end{aligned}$$

This implies the mean-square stability of the zero solution of (2.3).

For the asymptotic mean-square stability of the zero solution of (2.3), note that for any initial value $X_h^0 \in L^2(\Omega; V_h)$ it holds that $\lim_{j\to\infty} \mathbb{E}[X_h^j \otimes X_h^j] = 0$ if and only if

$$\lim_{j\to\infty} \|(\mathcal{S}_j \cdots \mathcal{S}_0)\, \mathbb{E}[X_h^0 \otimes X_h^0]\|_{V_h^{(2)}} = 0,$$

for which a sufficient condition is given by

$$\lim_{j\to\infty} \|\mathcal{S}_j \cdots \mathcal{S}_0\|_{L(V_h^{(2)})} = 0.$$

This finishes the proof of the theorem. □

In many applications the operators $(D_{\Delta t,h}^{\mathrm{stoch},j}, j \in \mathbb{N}_0)$ are also independent and identically distributed (iid for short):

Example 3.5. *Consider the one-dimensional geometric Brownian motion driven by a real-valued Brownian motion* $(\beta(t), t \ge 0)$

$$\mathrm{d}X(t) = \lambda X(t)\mathrm{d}t + \sigma X(t)\mathrm{d}\beta(t), \qquad t \ge 0,$$

with initial condition $X(0) = x_0 \in \mathbb{R}$ *and* $\lambda, \sigma \in \mathbb{R}$*. One way to approximate the solution of the geometric Brownian motion is given by the explicit Euler–Maruyama scheme defined by*

the recursion

$$X_{j+1} = X_j + \lambda \Delta t X_j + \sigma \Delta \beta^j X_j,$$

for $j \in \mathbb{N}_0$*, where* $\Delta\beta^j = \beta(t_{j+1}) - \beta(t_j)$*, or by using a Milstein scheme given by*

$$X_{j+1} = X_j + \lambda \Delta t X_j + \sigma \Delta \beta^j X_j + \frac{\sigma^2}{2}\left((\Delta\beta^j)^2 - \Delta t\right) X_j.$$

Then the deterministic operators in (2.3)

$$D^{\mathrm{det}}_{\Delta t,\mathit{EM}} = D^{\mathrm{det}}_{\Delta t,\mathit{Mil}} = 1 + \lambda \Delta t$$

are equal for both schemes, and the corresponding approximations of the stochastic integrals are given by

$$D^{\mathrm{stoch},j}_{\Delta t,\mathit{EM}} = \sigma \Delta \beta^j, \quad D^{\mathrm{stoch},j}_{\Delta t,\mathit{Mil}} = \sigma \Delta \beta^j + \frac{\sigma^2}{2}\left((\Delta\beta^j)^2 - \Delta t\right)$$

for $j \in \mathbb{N}_0$*. Thus, both families of stochastic approximation operators are* $\mathcal{F}$*-compatible and consist of identically distributed linear operators satisfying our assumption that* $D^{\mathrm{stoch},j}_{\Delta t,h}$ *is independent of the filtration* $\mathcal{F}_{t_j}$ *since it is admissible.*

For numerical approximations (2.3) based on families of iid linear operators $(D^{\mathrm{stoch},j}_{\Delta t,h}, j \in \mathbb{N}_0)$ the following corollary provides a necessary and sufficient condition for the asymptotic mean-square stability of the zero solution of (2.3) in terms of a condition on the spectrum of a single linear operator $\mathcal{S} \in L(V_h^{(2)})$.

Corollary 3.6. *Let* $X_h = (X_h^j, j \in \mathbb{N}_0)$ *be given by the approximation scheme* (2.3) *such that* $(D^{\mathrm{stoch},j}_{\Delta t,h}, j \in \mathbb{N}_0)$ *is iid. Then the zero solution of the discrete approximation scheme* (2.3) *is asymptotically mean-square stable if and only if*

$$\mathcal{S} = D^{\mathrm{det}}_{\Delta t,h} \otimes D^{\mathrm{det}}_{\Delta t,h} + \mathbb{E}[D^{\mathrm{stoch},0}_{\Delta t,h} \otimes D^{\mathrm{stoch},0}_{\Delta t,h}] \in L(V_h^{(2)})$$

satisfies $\rho(\mathcal{S}) < 1$*, where* $\rho(\mathcal{S})$ *is the spectral radius of* $\mathcal{S}$ *defined as* $\rho(\mathcal{S}) = \max_{i=1,\dots,N_h^2} |\lambda_i|$ *for* $\lambda_1, \dots, \lambda_{N_h^2}$ *being the eigenvalues of the operator* $\mathcal{S}$*.*

Furthermore, a sufficient condition for the zero solution of the discrete approximation scheme (2.3) *being asymptotically mean-square stable is given by* $\|\mathcal{S}\|_{L(V_h^{(2)})} < 1$*.*

Proof. First note that due to the identical distribution of the operators $(D^{\mathrm{stoch},j}_{\Delta t,h}, j \in \mathbb{N}_0)$, it is clear that

$$\mathbb{E}\left[D^{\mathrm{stoch},j}_{\Delta t,h} \otimes D^{\mathrm{stoch},j}_{\Delta t,h}\right] = \mathbb{E}\left[D^{\mathrm{stoch},0}_{\Delta t,h} \otimes D^{\mathrm{stoch},0}_{\Delta t,h}\right].$$

Hence, we get for all $j \in \mathbb{N}$, $\mathcal{S}_j = \mathcal{S}$ and thus, it holds that

$$\mathbb{E}[X_h^{j+1} \otimes X_h^{j+1}] = (\mathcal{S}_j \cdots \mathcal{S}_0)\, \mathbb{E}[X_h^0 \otimes X_h^0] = \mathcal{S}^{j+1}\, \mathbb{E}[X_h^0 \otimes X_h^0]$$

by the same arguments as in Theorem 3.4. As a consequence, $\lim_{j\to\infty}\mathbb{E}[X_h^j\otimes X_h^j]=0$ if and only if $\lim_{j\to\infty}\mathcal{S}^{j+1}=0$, which is equivalent to $\rho(\mathcal{S})<1$ by the same arguments as, e.g., in [7, 16, 10]. This completes the proof of the first statement.

Since $\rho(\mathcal{S})\leq\|\mathcal{S}\|_{L(V_h^{(2)})}$, a sufficient condition for the asymptotic mean-square stability of the zero solution of (2.3) is given by $\|\mathcal{S}\|_{L(V_h^{(2)})}<1$. □

Note that this corollary extends the results on mean-square stability of finite-dimensional linear systems of SDEs in [10] to the case of SPDE approximations on an operator-valued level of consideration. In [10] the proposed method to analyse the mean-square stability properties of finite-dimensional systems relies on a matrix eigenvalue problem. For SPDE approximations, this approach is not suitable, since the dimension of the considered eigenvalue problem increases heavily if we refine in space, i.e., for every considered refinement parameter $h>0$, we have to compute the spectral radius of an $(N_h^2\times N_h^2)$-matrix. For this reason, we perform, in what follows, mean-square stability analysis of SPDE approximations by just using properties of the involved operators.

4 Application to Galerkin methods

We continue by applying the previous results to the analysis of some classical numerical approximations of (2.1). The considered methods are based on a spatial discretization using Galerkin methods combined with different one-step time integration schemes such as one-step Euler–Maruyama or Milstein methods. In order to use the analytic framework of [19], we assume the operator $-A:\mathcal{D}(-A)\subset H\to H$ of (2.1) to be densely defined, self-adjoint, and positive definite with compact inverse.

This assumption implies that $-A$ has a non-decreasing sequence of positive eigenvalues $(\lambda_i, i\in\mathbb{N})$ for an orthonormal basis of eigenfunctions $(e_i, i\in\mathbb{N})$ in H. One defines fractional powers of $-A$ by the relation

$$(-A)^{r/2}e_i=\lambda_i^{r/2}e_i$$

for all $i\in\mathbb{N}$ and $r>0$. Then, for each $r>0$, $\dot{H}^r=\mathcal{D}((-A)^{r/2})$ with inner product

$$\langle\cdot,\cdot\rangle_r=\left\langle(-A)^{r/2}\cdot,(-A)^{r/2}\cdot\right\rangle_H$$

defines a separable Hilbert space. For details on the spaces $\dot{H}^r$, see, e.g., [19, Appendix B]. Let us further from here on assume that $G\in L(H;L(U;H))$.

For the spatial approximation of the solution of (2.1) we employ a Galerkin method. For this, let the sequence $(V_h, h\in(0,1])$ of finite-dimensional subspaces fulfil $V_h\subset\dot{H}^1\subset H$. The discrete operator $-A_h:V_h\to V_h$ is defined on each $v_h\in V_h$ by letting $-A_hv_h$ be the unique element of V_h such that

$$\langle -A_hv_h,w_h\rangle_H=\langle v_h,w_h\rangle_1=\left\langle(-A)^{1/2}v_h,(-A)^{1/2}w_h\right\rangle_H$$

for all $w_h \in V_h$. This definition implies that $-A_h$ is self-adjoint and positive definite on V_h and therefore has a sequence of orthonormal eigenfunctions $(e_{h,i}, i = 1, \dots, N_h)$ and positive non-decreasing eigenvalues $(\lambda_{h,i}, i = 1, \dots, N_h)$. By using basic properties of the Rayleigh quotient, we can bound the smallest eigenvalue $\lambda_{h,1}$ of $-A_h$ from below by the smallest eigenvalue λ_1 of $-A$ through

$$\lambda_{h,1} = \min_{v_h \in V_h \setminus \{0\}} \frac{\langle v_h, v_h \rangle_1}{\|v_h\|_H^2} \geq \min_{v \in H \setminus \{0\}} \frac{\langle v, v \rangle_1}{\|v\|_H^2} = \lambda_1, \tag{4.1}$$

since $V_h \subset H$, cf. [8]. This estimate turns out to be a useful inequality when investigating the connection between the asymptotic mean-square stability of the zero solution of an SPDE and its approximation.

Let us further assume in this section that the square-integrable martingale $M = L$ is a Lévy process and therefore has a stationary covariance $Q \in L(U)$ which is a self-adjoint and positive semidefinite operator of trace class. For the definition and properties of U-valued Lévy processes, the reader is referred to [26]. The assumed properties on Q imply by results in [26, Chapter 4] that there exists an orthonormal basis $(f_i, i \in \mathbb{N})$ of U and a non-increasing sequence of non-negative real numbers $(\mu_i, i \in \mathbb{N})$ such that for all $i \in \mathbb{N}$, $Qf_i = \mu_i f_i$ with $\operatorname{tr}(Q) = \sum_{i=1}^{\infty} \mu_i < \infty$ and that L admits a Karhunen–Loève expansion

$$L(t) = \sum_{i=1}^{\infty} \sqrt{\mu_i} L_i(t) f_i, \tag{4.2}$$

where $(L_i, i \in \mathbb{N})$ is a family of real-valued, square-integrable, uncorrelated Lévy processes fulfilling $\mathbb{E}[(L_i(t))^2] = t$ for all $t \geq 0$. Note that due to the martingale property of L, the real-valued Lévy processes satisfy $\mathbb{E}[L_i(t)] = 0$ for all $t \geq 0$ and $i \in \mathbb{N}$. This implies, together with the stationarity of the Lévy increments, that for all $i \in \mathbb{N}$ and $j \in \mathbb{N}_0$,

$$\mathbb{E}[\Delta L_i^j] = \mathbb{E}[\Delta L_1^0] = \mathbb{E}[L_1(\Delta t)] = 0,$$

where $\Delta L_i^j = L_i(t_{j+1}) - L_i(t_j)$.

Since the series representation of the Lévy process L can consist of infinitely many one-dimensional stochastic processes $L_i, i \in \mathbb{N}$, an additional approximation of L might be necessary for implementations of the fully discrete scheme (2.3). A common practice to approximate the U-valued process L is to truncate the corresponding Karhunen–Loève expansion, i.e., for $\kappa \in \mathbb{N}$, define the finite-dimensional approximation

$$L^{\kappa}(t) = \sum_{i=1}^{\kappa} \sqrt{\mu_i} L_i(t) f_i$$

of $L(t)$. Note that the truncation parameter κ has to be chosen appropriately such that the resulting truncation error does not dominate the approximation error of the numerical method. As it is noted in, e.g., [4, 6, 24], κ should be coupled with the convergence rate of the considered numerical scheme and the decay of the eigenvalues of Q.

Within this work, we consider the more general framework of numerical methods based on the original Karhunen–Loève expansion (4.2) of L. However, we emphasize that this does not restrict the applicability of the results since all of them can be easily adapted to the case when the approximated stochastic process $L^\kappa(t)$ is considered instead of $L(t)$ by choosing $\mu_k = 0$ for all $k > \kappa$.

We end this section by extending Example 2.1 with two explicit choices of the operator $G \in L(H; L(U; H))$.

Example 4.1. *Consider the setting of Example 2.1. It is known (see, e.g., [19, Chapter 6]) that the eigenvalues and eigenfunctions of the operator $-A$ are given by*

$$\lambda_i = \nu i^2 \pi^2, \quad e_i(y) = \sqrt{2}\sin(i\pi y), \qquad i \in \mathbb{N}, y \in [0,1].$$

We first assume, for simplicity, that $U = H = L^2([0,1])$ and that the operator Q diagonalizes with respect to the eigenbasis of $-A$, i.e., $f_i = e_i$ for all $i \in \mathbb{N}$. For this choice, we consider the operator $G = G_1$ that gives rise to a geometric Brownian motion in infinite dimensions*, cf. [19, Section 6.4]. It is for all $u, v \in H$ defined by the equation*

$$G_1(v)u = \sum_{i=1}^{\infty} \langle v, e_i\rangle_H \langle u, e_i\rangle_H e_i.$$

As a second example, we let $U = \dot{H}^1$ with the same diagonalization assumption as before, i.e., $f_i = \lambda_i^{1/2} e_i$ for all $i \in \mathbb{N}$. Here, we let the operator $G = G_2$ be a Nemytskii operator *which is defined pointwise for $x \in [0,1]$, $u \in \dot{H}^1$ and $v \in H$ by*

$$(G_2(v)u)[x] = v(x)u(x).$$

It is known that both choices of G are linear mappings from H to $L_{HS}(Q^{1/2}(U); H)$ (see, e.g., [19] and [21]) but we need to check that $G \in L(H; L(U; H))$.

To this end, note that for $u, v \in H$, by the triangle inequality and Cauchy–Schwarz we have for G_1

$$\|G_1(v)u\|_H \le \sum_{i=1}^{\infty} |\langle v, e_i\rangle_H||\langle u, e_i\rangle_H| \le \left(\sum_{i=1}^{\infty} \langle v, e_i\rangle_H^2\right)^{1/2} \left(\sum_{i=1}^{\infty} \langle u, e_i\rangle_H^2\right)^{1/2} = \|v\|_H \|u\|_H.$$

Next, for G_2 with $v \in H$ and $u \in \dot{H}^1$, it holds that

$$\begin{aligned}
\|G_2(v)u\|_H^2 &= \int_0^1 u(x)^2 v(x)^2 \,\mathrm{d}x = \int_0^1 \left(\sum_{i=1}^{\infty} \lambda_i^{1/2} \langle u, e_i\rangle_H \lambda_i^{-1/2} e_i(x)\right)^2 v(x)^2 \,\mathrm{d}x \\
&\le \left(\sum_{i=1}^{\infty} \lambda_i |\langle u, e_i\rangle_H|^2\right) \int_0^1 \left(\sum_{i=1}^{\infty} \lambda_i^{-1} e_i(x)^2\right) v(x)^2 \,\mathrm{d}x \\
&\le \|u\|_{\dot{H}^1}^2 \left(2\sum_{i=1}^{\infty} \lambda_i^{-1}\right) \int_0^1 v(x)^2 \,\mathrm{d}x = \left(2\sum_{i=1}^{\infty} \lambda_i^{-1}\right) \|u\|_{\dot{H}^1}^2 \|v\|_H^2.
\end{aligned}$$

Here, the first inequality is an application of the Cauchy–Schwarz inequality while the second follows from the fact that the sequence $(|e_i(x)|, i \in \mathbb{N})$ *is bounded by* $\sqrt{2}$ *for all* $x \in [0,1]$. *Therefore, we have*

$$\|G_1\|_{L(H;L(H))} \leq 1$$

and

$$\|G_2\|_{L(H;L(\dot{H}^1,H))} \leq \left(2\sum_{i=1}^{\infty} \lambda_i^{-1}\right)^{1/2}.$$

In the following sections, we investigate mean-square stability properties of fully discrete (Galerkin) approximations based on time discretizations with rational approximations.

4.1 Time discretization with rational approximations

Let us first recall that a *rational approximation of order* p of the exponential function is a rational function $R : \mathbb{C} \to \mathbb{C}$ satisfying that there exist constants $C, \delta > 0$ such that for all $z \in \mathbb{C}$ with $|z| < \delta$

$$|R(z) - \exp(z)| \leq C|z|^{p+1}.$$

Since R is rational there exist polynomials r_n and r_d such that for all $z \in \mathbb{C}$

$$R(z) = \frac{r_{\mathrm{n}}(z)}{r_{\mathrm{d}}(z)}.$$

We want to consider rational approximations of the semigroup S generated by the operator $-A$ and of its approximations $-A_h$ as they were considered in [29]. With the introduced notation, the linear operator $R(\Delta t A_h)$ is given for all $v_h \in V_h$ by

$$R(\Delta t A_h)v_h = r_{\mathrm{d}}^{-1}(\Delta t A_h) r_{\mathrm{n}}(\Delta t A_h)v_h = \sum_{k=1}^{N_h} \frac{r_{\mathrm{n}}(-\Delta t \lambda_{h,k})}{r_{\mathrm{d}}(-\Delta t \lambda_{h,k})} \langle v_h, e_{h,k}\rangle_H \, e_{h,k}. \tag{4.3}$$

We consider two cases of discretizations of the stochastic integral in combination with the rational approximation: the operator $D_{\Delta t,h}^{\mathrm{stoch},j}$ is first based on an Euler–Maruyama scheme and then on a Milstein scheme.

Let us start with the mean-square stability properties of a Galerkin Euler–Maruyama method, which is given by the recursion

$$\begin{aligned} X_h^{j+1} &= \left(R(\Delta t A_h) + r_{\mathrm{d}}^{-1}(\Delta t A_h)(\Delta t P_h F + P_h G(\cdot)\Delta L^j)\right) X_h^j, \\ X_h^0 &= P_h X_0 \end{aligned} \tag{4.4}$$

for $j \in \mathbb{N}_0$, where $\Delta L^j = L(t_{j+1}) - L(t_j)$. The corresponding operators from the fully

discrete scheme (2.3) are then given by

$$\begin{aligned} D^{\mathrm{det}}_{\Delta t,h} &= R(\Delta t A_h) + r_{\mathrm{d}}^{-1}(\Delta t A_h)\Delta t P_h F \\ D^{\mathrm{stoch},j}_{\Delta t,h} &= r_{\mathrm{d}}^{-1}(\Delta t A_h) P_h G(\cdot)\Delta L^j. \end{aligned} \tag{4.5}$$

Note that the linear operators $(D^{\mathrm{stoch},j}_{\Delta t,h}, j \in \mathbb{N}_0)$ satisfy all assumptions of Corollary 3.6 since they only depend on the Lévy increments $(\Delta L^j, j \in \mathbb{N}_0)$. For this type of numerical approximation, the result from Corollary 3.6 can be specified:

Proposition 4.2. *The zero solution of the numerical method* (4.4) *is asymptotically mean-square stable if and only if*

$$\mathcal{S} = D^{\mathrm{det}}_{\Delta t,h} \otimes D^{\mathrm{det}}_{\Delta t,h} + \Delta t\,(C \otimes C)q \in L(V_h^{(2)})$$

satisfies that $\rho(\mathcal{S}) < 1$*, where* $D^{\mathrm{det}}_{\Delta t,h}$ *is given in* (4.5)*,* $C \in L(U; L(V_h))$ *with*

$$Cu = r_{\mathrm{d}}^{-1}(\Delta t A_h) P_h G(\cdot)u,$$

and $q = \sum_{k=1}^{\infty} \mu_k f_k \otimes f_k \in U^{(2)}$*.*

Proof. First, note that since V_h is finite-dimensional, $L(V_h) = L_{\mathrm{HS}}(V_h)$ so $(C \otimes C)$ is well-defined as an element of $L(U^{(2)}, L^{(2)}_{\mathrm{HS}}(V_h)) \subset L(U^{(2)}, L(V_h^{(2)}))$ by [17, Lemma 3.1(ii)]. Hence, for every $j \in \mathbb{N}$, we have

$$\mathbb{E}[D^{\mathrm{stoch},j}_{\Delta t,h} \otimes D^{\mathrm{stoch},j}_{\Delta t,h}] = \mathbb{E}[C\Delta L^j \otimes C\Delta L^j] = (C \otimes C)\,\mathbb{E}[\Delta L^j \otimes \Delta L^j].$$

Since $\mathbb{E}[\Delta L^j \otimes \Delta L^j] = \Delta t\, q$ by Lemma A.1, the proof is completed with Corollary 3.6. □

The still rather abstract condition on mean-square stability is used to derive explicit sufficient conditions in the following corollary.

Corollary 4.3. *A sufficient condition for the asymptotic mean-square stability of the zero solution of* (4.4) *is then given by*

$$\begin{aligned} \Bigg(\max_{k=1,\dots,N_h} |R(-\Delta t\lambda_{h,k})| + \max_{k=1,\dots,N_h} |r_{\mathrm{d}}^{-1}(-\Delta t\lambda_{h,k})|\Delta t \|F\|_{L(H)}\Bigg)^2 \\ + \max_{k=1,\dots,N_h} |r_{\mathrm{d}}^{-1}(-\Delta t\lambda_{h,k})|^2 \Delta t\,\mathrm{tr}(Q)\|G\|^2_{L(H;L(U;H))} < 1. \end{aligned}$$

Proof. We first note that by the triangle inequality and the properties of the linear operator induced by the rational approximation R defined in Equation (4.3) we obtain that

$$\begin{aligned} \|D^{\mathrm{det}}_{\Delta t,h}\|_{L(V_h)} &= \|R(\Delta t A_h) + r_{\mathrm{d}}^{-1}(\Delta t A_h)\Delta t P_h F\|_{L(V_h)} \\ &\leq \max_{k=1,\dots,N_h} |R(-\Delta t\lambda_{h,k})| + \max_{k=1,\dots,N_h} |r_{\mathrm{d}}^{-1}(-\Delta t\lambda_{h,k})|\Delta t\|F\|_{L(H)} \end{aligned}$$

and, similarly

$$\begin{aligned}\|C\|_{L(U;L(V_h))} &\le \|r_{\mathrm{d}}^{-1}(\Delta t A_h)\|_{L(V_h)}\|G\|_{L(H;L(U;H))}\\ &\le \max_{k=1,\dots,N_h}|r_{\mathrm{d}}^{-1}(-\Delta t\lambda_{h,k})|\|G\|_{L(H;L(U;H))}.\end{aligned}$$

Since

$$\begin{aligned}\|\mathcal{S}\|_{L(V_h^{(2)})} &\le \|D_{\Delta t,h}^{\det}\otimes D_{\Delta t,h}^{\det}\|_{L(V_h^{(2)})} + \Delta t\|(C\otimes C)q\|_{L(V_h^{(2)})}\\ &\le \|D_{\Delta t,h}^{\det}\otimes D_{\Delta t,h}^{\det}\|_{L(V_h^{(2)})} + \Delta t\sum_{k=1}^{\infty}\mu_k\|(Cf_k\otimes Cf_k)\|_{L(V_h^{(2)})}\\ &\le \|D_{\Delta t,h}^{\det}\otimes D_{\Delta t,h}^{\det}\|_{L(V_h^{(2)})} + \Delta t\sum_{k=1}^{\infty}\mu_k\|Cf_k\|_{L(V_h)}^2\\ &\le \|D_{\Delta t,h}^{\det}\|_{L(V_h)}^2 + \Delta t\,\mathrm{tr}(Q)\|C\|_{L(U;L(V_h))}^2,\end{aligned}$$

we obtain the claimed condition, which is sufficient by Corollary 3.6. □

Having introduced sufficient conditions for Euler–Maruyama approximations, we continue with the Milstein scheme, which is of higher order than the Euler–Maruyama scheme. This is considered in [4] given by the recursion

$$\begin{aligned}X_h^{j+1} = \big(R(\Delta t A_h) + r_{\mathrm{d}}^{-1}(\Delta t A_h)(\Delta t P_h F + P_h G(\cdot)\Delta L^j)\big)X_h^j\\ + \int_{t_j}^{t_{j+1}} r_{\mathrm{d}}^{-1}(\Delta t A_h)P_h G\Big(\int_{t_j}^{s} G(X_h^j)\,\mathrm{d}L(r)\Big)\,\mathrm{d}L(s).\end{aligned} \tag{4.6}$$

The iterated stochastic integrals can be represented with the Karhunen–Loève expansion (4.2) of the Lévy process by

$$\begin{aligned}&\int_{t_j}^{t_{j+1}} r_{\mathrm{d}}^{-1}(\Delta t A_h)P_h G\Big(\int_{t_j}^{s} G(X_h^j)\,\mathrm{d}L(r)\Big)\,\mathrm{d}L(s)\\ &\qquad = \sum_{k,\ell=1}^{\infty}\sqrt{\mu_k\mu_\ell}\,r_{\mathrm{d}}^{-1}(\Delta t A_h)P_h G(G(X_h^j)f_k)f_\ell\int_{t_j}^{t_{j+1}}\int_{t_j}^{s}\mathrm{d}L_k(r)\,\mathrm{d}L_\ell(s).\end{aligned}$$

Thus, the stochastic operator $D_{\Delta t,h}^{\mathrm{stoch},j}$ can be written as the sum of the stochastic operator of the Euler–Maruyama scheme (4.5) and an operator including the iterated stochastic integrals, i.e.,

$$D_{\Delta t,h}^{\mathrm{stoch},j} = D_{\Delta t,h}^{\mathrm{EM},j} + D_{\Delta t,h}^{\mathrm{M},j}, \tag{4.7}$$

where

$$\begin{aligned}D_{\Delta t,h}^{\mathrm{EM},j} &= r_{\mathrm{d}}^{-1}(\Delta t A_h)P_h G(\cdot)\Delta L^j,\\ D_{\Delta t,h}^{\mathrm{M},j} &= \sum_{k,\ell=1}^{\infty} r_{\mathrm{d}}^{-1}(\Delta t A_h)\sqrt{\mu_k\mu_\ell}P_h G(G(\cdot)f_k)f_\ell\int_{t_j}^{t_{j+1}}\int_{t_j}^{s}\mathrm{d}L_k(r)\,\mathrm{d}L_\ell(s).\end{aligned}$$

Remark 4.4. In order to compute the iterated integrals of $D^{\mathrm{M},j}_{\Delta t,h}$, one may assume (cf. [4, 15]) that for all H-valued, $\mathcal{F}_t$-adapted stochastic processes $\chi = (\chi(t), t \geq 0)$ and all $i, j \in \mathbb{N}$, the diffusion operator G satisfies the commutativity condition

$$G(G(\chi)f_j)f_i = G(G(\chi)f_i)f_j.$$

Under this assumption, which is satisfied for the operators of Example 4.1, the iterated integrals can be written in terms of the Lévy increments, i.e.,

$$\begin{aligned}\int_{t_j}^{t_{j+1}} r_{\mathrm{d}}^{-1}(\Delta t A_h)P_h G\left(\int_{t_j}^{s} G(X_h^j)\mathrm{d}L(r)\right)\mathrm{d}L(s)\\ = \frac{1}{2}\sum_{k,\ell=1}^{\infty}\sqrt{\mu_k\mu_\ell}\, r_{\mathrm{d}}^{-1}(\Delta t A_h)P_h G(G(X_h^j)f_k)f_\ell(\Delta L_k^j \Delta L_\ell^j - \Delta[L_k, L_\ell]^j),\end{aligned}$$

where $\Delta[L_k, L_\ell]^j = [L_k, L_\ell]_{t_{j+1}} - [L_k, L_\ell]_{t_j}$. Here, $[L_k, L_\ell]_t$ is the quadratic covariation of L_k and L_ℓ evaluated at $t \geq 0$, which is straightforward to compute when L_k, L_ℓ are jump-diffusion processes (cf. [4]). For the simulation of more general Lévy processes in the context of SPDE approximation, we refer to [12].

If the diffusion operator does not satisfy this commutativity condition, the simulation of the iterated integral cannot be done using only the increments of the Lévy processes $(L_i, i \in \mathbb{N})$. In case that L is a Q-Wiener process, there are ways for the simulation of the involved iterated integrals which can be found, e.g., in [30]. A mean-square stability analysis of Milstein schemes for finite-dimensional SDEs driven by standard Wiener processes has been done for non-commuting diffusion matrices in [10], where the authors used appropriately truncated stochastic Lévy areas.

As for the Euler–Maruyama scheme of Proposition 4.2, this Milstein scheme allows the result from Corollary 3.6 to be specified in the following proposition.

Proposition 4.5. *Assume that the bilinear mapping*

$$C'(u_1, u_2) = (r_{\mathrm{d}}^{-1}(\Delta t A_h)P_h G(G(\cdot)u_1)u_2)$$

for $u_1, u_2 \in U$ can be uniquely extended to a mapping $C' \in L(U^{(2)}, L(V_h))$. Then the zero solution of the numerical method (4.6) *is asymptotically mean-square stable if and only if*

$$\mathcal{S} = D^{\mathrm{det}}_{\Delta t,h} \otimes D^{\mathrm{det}}_{\Delta t,h} + \Delta t\,(C \otimes C)q + \frac{\Delta t^2}{2}(C' \otimes C')q'$$

satisfies that $\rho(\mathcal{S}) < 1$. Here, $q' = \sum_{k,\ell=1}^{\infty}\mu_k\mu_\ell(f_k \otimes f_\ell) \otimes (f_k \otimes f_\ell) \in U^{(4)}$ while C and q are defined as in Proposition 4.2.

Proof. Note that $C' \otimes C' : U^{(4)} \to L(V_h^{(2)})$ and $C' \otimes C : U^{(2)} \otimes U \to L(V_h^{(2)})$ are well-defined by the same arguments as in Proposition 4.2. By Equation (4.7), we obtain for

$j \in \mathbb{N}_0$

$$\mathbb{E}[D_{\Delta t,h}^{\mathrm{stoch},j} \otimes D_{\Delta t,h}^{\mathrm{stoch},j}] = \mathbb{E}[D_{\Delta t,h}^{\mathrm{EM},j} \otimes D_{\Delta t,h}^{\mathrm{EM},j}] + \mathbb{E}[D_{\Delta t,h}^{\mathrm{M},j} \otimes D_{\Delta t,h}^{\mathrm{EM},j}] + \mathbb{E}[D_{\Delta t,h}^{\mathrm{EM},j} \otimes D_{\Delta t,h}^{\mathrm{M},j}] + \mathbb{E}[D_{\Delta t,h}^{\mathrm{M},j} \otimes D_{\Delta t,h}^{\mathrm{M},j}].$$

For the first term, observe that the underlying operator $D_{\Delta t,h}^{\mathrm{EM},j}$ coincides with the operator $D_{\Delta t,h}^{\mathrm{stoch},j}$ from Equation (4.5) and therefore, the first two components of $\mathcal{S}$ follow from the proof of Proposition 4.2. Writing $\Delta^{(2)} L = \sum_{k,\ell=1}^{\infty} \sqrt{\mu_k \mu_\ell} \left(\int_{t_j}^{t_{j+1}} \int_{t_j}^{s} \mathrm{d}L_k(r)\, \mathrm{d}L_\ell(s) \right) f_k \otimes f_\ell$, we get for the second term

$$\mathbb{E}[D_{\Delta t,h}^{\mathrm{M},j} \otimes D_{\Delta t,h}^{\mathrm{EM},j}] = \mathbb{E}\left[C' \Delta^{(2)} L^j \otimes C \Delta L^j \right] = (C' \otimes C)\, \mathbb{E}\left[\Delta^{(2)} L^j \otimes \Delta L^j \right]$$

and by Lemma A.2 $\mathbb{E}\left[\Delta^{(2)} L^j \otimes \Delta L^j \right] = 0$. Analogous calculations show that the third term is equal to zero as well. Finally, we have

$$\mathbb{E}[D_{\Delta t,h}^{\mathrm{M},j} \otimes D_{\Delta t,h}^{\mathrm{M},j}] = \mathbb{E}\left[C' \Delta^{(2)} L^j \otimes C' \Delta^{(2)} L^j \right] = (C' \otimes C')\, \mathbb{E}\left[\Delta^{(2)} L^j \otimes \Delta^{(2)} L^j \right]$$

and by Lemma A.2

$$E\left[\Delta^{(2)} L^j \otimes \Delta^{(2)} L^j \right] = \frac{\Delta t^2}{2} q'.$$

The statement now follows directly from Corollary 3.6. □

Before we apply the derived conditions to examples of rational approximations, we finish this part with a remark on the regularity assumption on the Milstein term.

Remark 4.6. The assumption in Proposition 4.5, that the bilinear mapping C' can be uniquely extended to a mapping in the space $L(U^{(2)}; L(V_h))$, holds for the operators G_1 and G_2 in the setting of Example 4.1. One can get rid of this assumption by using that the bound on $G \in L(H; L(U; H))$ allows for an extension of the bilinear mapping to the *projective tensor product space* $U \otimes_\pi U$, cf. [17]. One would then have to assume additional regularity on L to ensure that $\Delta^{(2)} L^j$ in the proof of Proposition 4.5 is in the space $L^2(\Omega; U \otimes_\pi U)$.

4.2 Examples of rational approximations

Let us next consider specific choices of rational approximations R and investigate their influence on the mean-square stability properties of approximations of SPDEs driven by Lévy noise. First, we derive sufficient conditions based on Corollary 4.3 for the asymptotic mean-square stability of the zero solution of Euler–Maruyama schemes with standard rational approximations for fixed time step size Δt and spatial refinement parameter h. More specifically, we consider the backward Euler, the Crank–Nicolson, and the forward Euler scheme.

Theorem 4.7. *Consider the approximation scheme* (4.4).

1 (Backward Euler scheme) *Let R be given by*

$$R(z) = \frac{1}{1-z}.$$

Then a sufficient condition for the asymptotic mean-square stability of the zero solution is

$$\frac{(1+\Delta t\|F\|_{L(H)})^2 + \Delta t\,\mathrm{tr}(Q)\|G\|^2_{L(H;L(U;H))}}{(1+\Delta t\lambda_{h,1})^2} < 1.$$

2 (Crank–Nicolson scheme) *Let R be given by*

$$R(z) = \frac{1+z/2}{1-z/2}.$$

Then a sufficient condition for the asymptotic mean-square stability of the zero solution is

$$\left(\max_{k\in\{1,N_h\}}\left|\frac{1-\Delta t\lambda_{h,k}/2}{1+\Delta t\lambda_{h,k}/2}\right| + \Delta t\frac{\|F\|_{L(H)}}{(1+\Delta t\lambda_{h,1}/2)}\right)^2 + \Delta t\frac{\mathrm{tr}(Q)\|G\|^2_{L(H;L(U;H))}}{(1+\Delta t\lambda_{h,1}/2)^2} < 1.$$

3 (Forward Euler scheme) *Let R be given by*

$$R(z) = 1+z.$$

Then a sufficient condition for the asymptotic mean-square stability of the zero solution is

$$\left(\max_{\ell\in\{1,N_h\}}|1-\Delta t\lambda_{h,\ell}| + \Delta t\|F\|_{L(H)}\right)^2 + \Delta t\,\mathrm{tr}(Q)\|G\|^2_{L(H;L(U;H))} < 1.$$

Proof. Let us start with the backward Euler scheme. Since the functions $r_{\mathrm{d}}^{-1}(z)$ and $R(z)$ are equal and it holds for all $k = 1,\dots,N_h$ that $|R(-\Delta t\lambda_{h,k})| \le |R(-\Delta t\lambda_{h,1})|$, we obtain by Corollary 4.3 asymptotic mean-square stability if

$$(1+\Delta t\lambda_{h,1})^{-2}\left((1+\Delta t\|F\|_{L(H)})^2 + \Delta t\,\mathrm{tr}(Q)\|G\|^2_{L(H;L(U;H))}\right) < 1.$$

For the Crank–Nicolson scheme, note that R is decreasing on $\mathbb{R}^-$ and that $R(z) \in [-1,1]$ for all $z \in \mathbb{R}^-$. Thus, the maximizing eigenvalue is either the largest, λ_{h,N_h}, or the smallest, $\lambda_{h,1}$, and therefore,

$$|R(-\Delta t\lambda_{h,k})| \le \max_{\ell\in\{1,N_h\}}|R(-\Delta t\lambda_{h,\ell})|.$$

Since $|r_{\mathrm{d}}^{-1}(-\Delta t\lambda_{h,k})| \le |r_{\mathrm{d}}^{-1}(-\Delta t\lambda_{h,1})|$ for all $k = 1,\dots,N_h$, the claim follows with Corollary 4.3.

By the same arguments, we obtain for the forward Euler scheme that $|R(-\Delta t\lambda_{h,i})|$ is

maximized either at $z = -\Delta t\lambda_{h,1}$ or $z = -\Delta t\lambda_{h,N_h}$. Therefore, since $r_{\mathrm{d}}^{-1}(z) = 1$, the claim follows again with Corollary 4.3, which finishes the proof. □

Remark 4.8. In the homogeneous case, i.e., $F = 0$, the stability condition in Theorem 4.7(1) reduces to

$$\operatorname{tr}(Q)\|G\|^2_{L(H;L(U;H))} < \lambda_{h,1}(2 + \Delta t\lambda_{h,1})$$

so that even if the zero solution of (2.1) is asymptotically mean-square unstable, the zero solution of its approximation (4.4) can always be rendered stable by letting Δt be large enough.

Finally, for the Milstein scheme, Proposition 4.5 allows us to show the following sufficient condition for asymptotic mean-square stability.

Proposition 4.9. *Let $D^{\mathrm{det}}_{\Delta t,h}$ be given by Equation* (4.5) *and let $D^{\mathrm{stoch},j}_{\Delta t,h}$ be defined by Equation* (4.7)*. Consider the backward Euler approximation with $R(z) = (1 - z)^{-1}$. Then, under the same assumptions as in Proposition 4.5, a sufficient condition for the asymptotic mean-square stability of the zero solution of* (4.6) *is*

$$(1 + \Delta t\|F\|_{L(H)})^2 + \Delta t \operatorname{tr}(Q)\|G\|^2_{L(H;L(U;H))} + \frac{\Delta t^2}{2}\operatorname{tr}(Q)^2\|G\|^4_{L(H;L(U;H))} < (1 + \Delta t\lambda_{h,1})^2.$$

Proof. In the same way as in the proof of Corollary 4.3, we bound

$$\|(C' \otimes C')q'\|_{L(V_h^{(2)})} \le \|C'\|^2_{L(U^{(2)};L(V_h))} \operatorname{tr}(Q)^2 \le (1 + \Delta t\lambda_{h,1})^{-2}\|G\|^4_{L(H;L(U;H))} \operatorname{tr}(Q)^2.$$

Hence, our assumption ensures that $\|\mathcal{S}\|_{L(V_h^{(2)})} < 1$, which by Corollary 3.6 is a sufficient condition for the asymptotic mean-square stability of the zero solution of (4.6). □

Note that the sufficient condition for the asymptotic mean-square stability of the zero solution derived here for the Milstein scheme is more restrictive than the sufficient condition presented in Theorem 4.7(1) for the backward Euler–Maruyama method due to the additional positive term in the estimate in Proposition 4.9.

4.3 Relation to the mild solution

To be able to connect existing results on the asymptotic mean-square stability of the zero solution of (2.1) to our results for the discrete scheme (2.3) outlined in Section 4.2, we now restrict ourselves to the case that L is a Q-Wiener process denoted by $W = (W(t), t \ge 0)$, i.e., for $t \in \mathbb{R}^+$ we consider the SDE

$$\mathrm{d}X(t) = (AX(t) + F(X(t)))\mathrm{d}t + G(X(t))\mathrm{d}W(t). \tag{4.8}$$

This restriction is caused by the available literature for the analytical solution. Note that for Q-Wiener processes, the Karhunen–Loève expansion (4.2) becomes

$$W(t) = \sum_{i=1}^{\infty} \sqrt{\mu_i}\beta_i(t) f_i,$$

where $(\beta_i, i \in \mathbb{N})$ is a sequence of independent, real-valued Brownian motions.

The following result, which is a special case of [23, Proposition 3.1.1], provides a sufficient condition for the asymptotic mean-square stability of the zero solution of (2.1) by a Lyapunov functional approach.

Theorem 4.10. *Assume that the initial value $X_0 = x_0 \in \dot{H}^1$ is deterministic and that there exists a constant $c > 0$ such that for all $v \in \dot{H}^2$ it holds that*

$$2\langle v, Av + F(v)\rangle_H + \operatorname{tr}[G(v)Q(G(v))^*] \leq -c\|v\|_H^2.$$

Then the zero solution of (4.8) *is asymptotically mean-square stable.*

We use this theorem to derive a sufficient condition for the simultaneous mean-square stability of the zero solutions of (4.8) and of the backward Euler approximation scheme (4.4).

Corollary 4.11. *Consider the backward Euler scheme in the framework of Theorem 4.7 and assume that $X_0 = x_0 \in \dot{H}^1$ is deterministic. Then*

$$2\left(\|F\|_{L(H)} - \lambda_1\right) + \operatorname{tr}(Q)\|G\|^2_{L(H;L(U;H))} < 0 \tag{4.9}$$

is a sufficient condition for the asymptotic mean-square stability of the zero solutions of both, (4.8) *and its approximation* (4.4)*, independent of the values of h and Δt.*

Proof. Let us assume that

$$2\left(\|F\|_{L(H)} - \lambda_1\right) + \operatorname{tr}(Q)\|G\|^2_{L(H;L(U;H))} < 0.$$

We first show that under this assumption the zero solution of (4.8) is asymptotically mean-square stable, which by Theorem 4.10 follows from

$$2\langle v, Av + F(v)\rangle_H + \operatorname{tr}[G(v)Q(G(v))^*] < -c\|v\|_H^2$$

for some $c > 0$. For the second term of this expression, note that for any $v \in \dot{H}^2$,

$$\begin{aligned}\operatorname{tr}[G(v)Q(G(v))^*] &= \operatorname{tr}[(G(v))^* G(v)Q] = \sum_{k=1}^{\infty} \langle G(v)Qf_k, G(v)f_k\rangle \\ &\leq \sum_{k=1}^{\infty} \mu_k \|G\|^2_{L(H;L(U;H))}\|v\|_H^2\|f_k\|_U^2 = \operatorname{tr}(Q)\|G\|^2_{L(H;L(U;H))}\|v\|_H^2,\end{aligned}$$

where the first equality follows from the fact that the trace operator is invariant under cyclic

permutations. Next, for the first term, we have that

$$\langle v, Av + F(v)\rangle = \langle v, F(v)\rangle + \langle v, Av\rangle \leq \|F\|_{L(H)}\|v\|_H^2 - \|v\|_1^2 \leq (\|F\|_{L(H)} - \lambda_1)\|v\|_H^2,$$

where the last inequality follows from the definition of $\|\cdot\|_1$. Altogether, we therefore find

$$\begin{aligned} &2\langle v, Av + F(v)\rangle_H + \operatorname{tr}[G(v)Q(G(v))^*] \\ &\qquad \leq \left(2\left(\|F\|_{L(H)} - \lambda_1\right) + \operatorname{tr}(Q)\|G\|_{L(H;L(U;H))}^2\right)\|v\|_H^2, \end{aligned}$$

which finishes the proof of the asymptotic mean-square stability of the zero solution of (4.8) with

$$c = -\left(2\left(\|F\|_{L(H)} - \lambda_1\right) + \operatorname{tr}(Q)\|G\|_{L(H;L(U;H))}^2\right).$$

In a second step, we show that under the same assumption the zero solution of (4.4) is asymptotically mean-square stable. We note that by (4.1) $\lambda_{h,1} \geq \lambda_1$. By Theorem 4.7, the asymptotic mean-square stability then follows if we can show that

$$(1 + \Delta t\|F\|_{L(H)})^2 + \Delta t \operatorname{tr}(Q)\|G\|_{L(H;L(U;H))}^2 < (1 + \Delta t\,\lambda_1)^2.$$

This inequality can be rewritten as

$$\Delta t\left(2\left(\|F\|_{L(H)} - \lambda_1\right) + \operatorname{tr}(Q)\|G\|_{L(H;L(U;H))}^2\right) + \Delta t^2\left(\|F\|_{L(H)}^2 - \lambda_1^2\right) < 0,$$

which is seen to hold by observing that the first term is negative by assumption and so is the second term, since

$$\begin{aligned} &\|F\|_{L(H)}^2 - \lambda_1^2 \\ &\quad = \left(\|F\|_{L(H)} + \lambda_1\right)\left(\|F\|_{L(H)} - \lambda_1\right) \\ &\quad \leq \left(\|F\|_{L(H)} + \lambda_1\right)\left(\left(\|F\|_{L(H)} - \lambda_1\right) + \frac{1}{2}\operatorname{tr}(Q)\|G\|_{L(H;L(U;H))}^2\right) < 0. \end{aligned}$$

This finishes the proof. □

Note that by Corollary 4.11 we get that whenever (4.9) is fulfilled, then, the backward Euler–Maruyama scheme preserves the qualitative behaviour of the analytical solution without any restriction on the refinement parameters h and Δt. Hence, under the condition that (4.9) is fulfilled, the backward Euler–Maruyama scheme can be applied to numerical methods that use different refinement parameters in parallel, since the zero solution of the numerical scheme (4.4) cannot become asymptotically mean-square unstable for any choice of h and Δt. An important example of such a method using a hierarchy of different refinement levels is the multilevel Monte Carlo estimator. It approximates efficiently a quantity of interest $\mathbb{E}[\varphi(X(T))]$ of the solution of (2.1) (see, e.g., [3, 5] for details). Here, a method preserving the qualitative behaviour of the analytical solution is required on every refinement

level, see, e.g., [1].

Remark 4.12. Based on Theorem 4.7, it is also possible to examine the relation between the asymptotic mean-square stability of the zero solution of (4.8) and its approximation by numerical schemes based on other rational approximations such as, e.g., the Crank–Nicolson or the forward Euler scheme. However, due to the nature of the sufficient conditions of Theorem 4.7 for these rational approximations, similar results for simultaneous mean-square stability of the zero solution as the one presented for the backward Euler scheme in Corollary 4.3 would include restrictions on the refinement parameters h and Δt.

For the Milstein scheme considered in Proposition 4.9 we can also derive a sufficient condition (independent of h and Δt) for the simultaneous mean-square stability of the zero solutions of (4.8) and of the approximation scheme (4.6). However, due to the additional term in the sufficient condition, the sufficient condition for the simultaneous mean-square stability is slightly more restrictive than in Corollary 4.11. More precisely we obtain the following:

Corollary 4.13. *Consider the backward Euler–Milstein scheme in the framework of Proposition 4.9 and assume further that $X_0 = x_0 \in \dot{H}^1$ is deterministic and $F = 0$. Then the inequality*

$$-\sqrt{2}\lambda_1 + \operatorname{tr}(Q)\|G\|^2_{L(H;L(U;H))} < 0$$

is a sufficient condition for the asymptotic mean-square stability of the zero solutions of both, (4.8) *and its approximation* (4.6)*, independent of h and Δt.*

Proof. Let us assume that $-\sqrt{2}\lambda_1 + \operatorname{tr}(Q)\|G\|^2_{L(H;L(U;H))} < 0$. Since

$$-2\lambda_1 + \operatorname{tr}(Q)\|G\|^2_{L(H;L(U;H))} < -\sqrt{2}\lambda_1 + \operatorname{tr}(Q)\|G\|^2_{L(H;L(U;H))} < 0,$$

we get the asymptotic mean-square stability of the zero solution of (4.8) by Corollary 4.11. The sufficient condition of Proposition 4.9 for the asymptotic mean-square stability of the zero solution of (4.6) can be rewritten as

$$\Delta t(-2\lambda_{h,1} + \operatorname{tr}(Q)\|G\|^2_{L(H;L(U;H))}) + \Delta t^2(-2\lambda^2_{h,1} + \operatorname{tr}(Q)^2\|G\|^4_{L(H;L(U;H))}) < 0.$$

The first part of this condition on the left hand side is negative since

$$-2\lambda_{h,1} + \operatorname{tr}(Q)\|G\|^2_{L(H;L(U;H))} < -\sqrt{2}\lambda_1 + \operatorname{tr}(Q)\|G\|^2_{L(H;L(U;H))} < 0.$$

For the second part note that $\sqrt{2}\lambda_1 > \operatorname{tr}(Q)\|G\|^2_{L(H;L(U;H))}$ by assumption and as a consequence it holds that

$$-2\lambda^2_{h,1} + \operatorname{tr}(Q)^2\|G\|^4_{L(H;L(U;H))} < 0.$$

Thus, we obtain the asymptotic mean-square stability of the zero solution of (4.6) by Propo-

sition 4.9, which completes the proof. □

5 Simulations

In this section we adopt the setting of Examples 2.1 and 4.1 and use numerical simulations to illustrate our theoretical results. We recall that we consider the stochastic heat equation on the spatial domain $[0,1]$ with homogeneous zero Dirichlet boundary conditions, i.e., for fixed $\nu > 0$ consider for $t \geq 0$

$$\mathrm{d}X(t) = \nu \Delta X(t)\,\mathrm{d}t + G(X(t))\,\mathrm{d}W(t). \tag{5.1}$$

We set the initial condition $X_0(x) = cx(1-x)$, where the constant $c = \sqrt{30}$ is chosen such that $\mathbb{E}[\|X_0\|_H^2] = 1$.

The eigenvalues $(\mu_i, i \in \mathbb{N})$ of the operator Q obtained by the relation $Qe_i = \mu_i e_i$ are assumed to be $\mu_i = C_\mu i^{-\alpha}$, where $C_\mu > 0$ and $\alpha > 1$. Here, C_μ can be used for scaling the noise intensity and α controls the regularity of the Q-Wiener process, see, e.g., [24, 22].

5.1 Spectral Galerkin methods

We examine spectral Galerkin methods for the stochastic heat equation with diffusion operator G_1 of Example 4.1 first. Following the approach presented in [19, Section 6.4], we obtain the infinite-dimensional counterpart of the geometric Brownian motion, because the mild solution of SPDE (5.1) can be represented by

$$X(t) = \sum_{i=1}^{\infty} \langle X(t), e_i \rangle_H e_i = \sum_{i=1}^{\infty} x_i(t) e_i,$$

where each of the coefficients $x_i(t)$ in the series representation of the mild solution corresponds to the solution of the one-dimensional geometric Brownian motion

$$\mathrm{d}x_i(t) = -\lambda_i x_i(t)\,\mathrm{d}t + \sqrt{\mu_i} x_i(t)\,\mathrm{d}\beta_i(t).$$

Furthermore, the second moment is given by

$$\mathbb{E}[\|X(T)\|_H^2] = \sum_{i=1}^{\infty} \mathbb{E}[|x_i(T)|^2] = \sum_{i=1}^{\infty} \langle X_0, e_i \rangle_H^2 \exp((-2\lambda_i + \mu_i)T).$$

Consequently, the zero solution of (5.1) is asymptotically mean-square stable if and only if for all $i \in \mathbb{N}$ it holds that $-2\lambda_i + \mu_i < 0$. By using the explicit representation of the eigenvalues λ_i and μ_i, this corresponds to the condition that for all $i \in \mathbb{N}$, it holds that $-2\nu i^2 \pi^2 + C_\mu i^{-\alpha} < 0$, which is equivalent to $-2\lambda_1 + \mu_1 = -2\nu\pi^2 + C_\mu < 0$. In reverse, the zero solution of (5.1) is asymptotically mean-square unstable if and only if $C_\mu > 2\nu\pi^2$.

For the spectral Galerkin approximation, we choose $V_h = \operatorname{span}(e_1, \ldots, e_{N_h})$, $N_h < \infty$. Thus, we want to find $X_h(t) = \sum_{k=1}^{N_h} \langle X(t), e_k \rangle_H e_k = \sum_{k=1}^{N_h} x_k(t) e_k$ of the semi-discretized

stochastic heat equation

$$\mathrm{d}X_h(t) = A_h X_h(t)\,\mathrm{d}t + P_h G_1(X_h(t))\,\mathrm{d}W(t). \tag{5.2}$$

To obtain a fully discrete scheme, we approximate the one-dimensional geometric Brownian motions in time by the three considered types of rational approximations: backward Euler, Crank–Nicolson, and forward Euler. Propositions 4.2 and 4.5 yield that the zero solution of the fully discrete approximation of (5.2) is asymptotically mean-square stable if and only if the corresponding linear operator $\mathcal{S}$ satisfies $\rho(\mathcal{S}) < 1$.

For computing the spectrum of the linear operator $\mathcal{S}$ from Proposition 4.2 (Euler–Maruyama scheme), we consider for $k, \ell = 1, \dots, N_h$

$$\begin{aligned}\mathcal{S}(e_k \otimes e_\ell) &= (D^{\det}_{\Delta t,h} \otimes D^{\det}_{\Delta t,h})(e_k \otimes e_\ell) + \Delta t\,((C \otimes C)q)\,(e_k \otimes e_\ell) \\ &= (D^{\det}_{\Delta t,h} e_k \otimes D^{\det}_{\Delta t,h} e_\ell) + \Delta t \sum_{m=1}^{\infty} \mu_m \Big(\big((Ce_m)e_k\big) \otimes \big((Ce_m)e_\ell\big) \Big).\end{aligned}$$

Since

$$D^{\det}_{\Delta t,h} e_k = R(\Delta t A_h)e_k = \sum_{r=1}^{N_h} R(-\Delta t \lambda_r)\langle e_k, e_r\rangle_H e_r = R(-\Delta t\lambda_k)e_k$$

and

$$\begin{aligned}(Ce_m)e_k &= r_{\mathrm{d}}^{-1}(\Delta t A_h)P_h G_1(e_k)e_m = r_{\mathrm{d}}^{-1}(\Delta t A_h)P_h \left(\sum_{n=1}^{\infty} \langle e_k, e_n\rangle_H \langle e_m, e_n\rangle_H e_n \right) \\ &= \delta_{k,m} r_{\mathrm{d}}^{-1}(\Delta t A_h)e_k = \delta_{k,m} r_{\mathrm{d}}^{-1}(-\Delta t\lambda_k)e_k,\end{aligned}$$

the eigenvalues $\Lambda_{k,\ell}$ corresponding to the eigenfunctions $e_k \otimes e_\ell$ are given by

$$\Lambda_{k,\ell} = R(-\Delta t\lambda_k)R(-\Delta t\lambda_\ell) + \delta_{k,\ell}\,\Delta t \mu_k\, r_{\mathrm{d}}^{-1}(-\Delta t\lambda_k) r_{\mathrm{d}}^{-1}(-\Delta t\lambda_\ell).$$

As another approximation we approximate the one-dimensional geometric Brownian motions by a Milstein scheme. For the linear operator $\mathcal{S}$ from Proposition 4.5 (Milstein scheme), the commutativity assumption on G of Remark 4.4 is fulfilled, and by using that $\Delta[\beta_k, \beta_\ell]^j = \delta_{k,\ell}\Delta t$ along with similar computations to those of the Euler–Maruyama scheme, one can show that the corresponding eigenvalues $\Lambda_{k,\ell}$ are given by

$$\Lambda_{k,\ell} = R(-\Delta t\lambda_k)R(-\Delta t\lambda_\ell) + \delta_{k,\ell}\, r_{\mathrm{d}}^{-1}(-\Delta t\lambda_k) r_{\mathrm{d}}^{-1}(-\Delta t\lambda_\ell) \left(\Delta t\mu_k + \frac{\Delta t^2 \mu_k^2}{2} \right). \tag{5.3}$$

The additional quadratic term comes from applying the linear operator $(C' \otimes C')q'$ to $e_k \otimes e_\ell$. Note that for both operators $\mathcal{S}$, the eigenvalues $\Lambda_{k,\ell}$ with $k \neq \ell$ satisfy

$$|\Lambda_{k,\ell}| = |R(-\Delta t\lambda_k)R(-\Delta t\lambda_\ell)| \leq R(-\Delta t\lambda_s)^2 \leq \Lambda_{s,s},$$

where $|R(-\Delta t\lambda_s)| = \max_{j=1,\dots,N_h} |R(-\Delta t\lambda_j)|$. Hence, $\rho(\mathcal{S}) < 1$ is equivalent to $|\Lambda_{k,k}| < 1$ for all $k = 1, \dots, N_h$. In Table 1 the eigenvalues $\Lambda_{k,k}$ and sufficient and necessary conditions for the asymptotic mean-square stability of the zero solution of (4.4) and (4.6) are presented for the considered rational approximations (backward Euler, Crank–Nicolson, and forward Euler).

rational approximation/ stochastic approximation	$\Lambda_{k,k}$	$\rho(\mathcal{S}) < 1 \Leftrightarrow$ for all $k = 1, \dots, N_h$:
backward Euler/EM	$\frac{1+\Delta t\mu_k}{(1+\Delta t\lambda_k)^2}$	$-2\lambda_k + \mu_k - \Delta t\lambda_k^2 < 0$
backward Euler/Milstein	$\frac{1+\Delta t\mu_k+\Delta t^2\mu_k^2/2}{(1+\Delta t\lambda_k)^2}$	$-2\lambda_k + \mu_k + \Delta t(-\lambda_k^2 + \mu_k^2/2) < 0$
Crank–Nicolson/EM	$\frac{(1-\Delta t\lambda_k/2)^2+\mu_k\Delta t}{(1+\Delta t\lambda_k/2)^2}$	$-2\lambda_k + \mu_k < 0$
forward Euler/EM	$(1 - \Delta t\lambda_k)^2 + \mu_k\Delta t$	$-2\lambda_k + \mu_k + \Delta t\lambda_k^2 < 0$

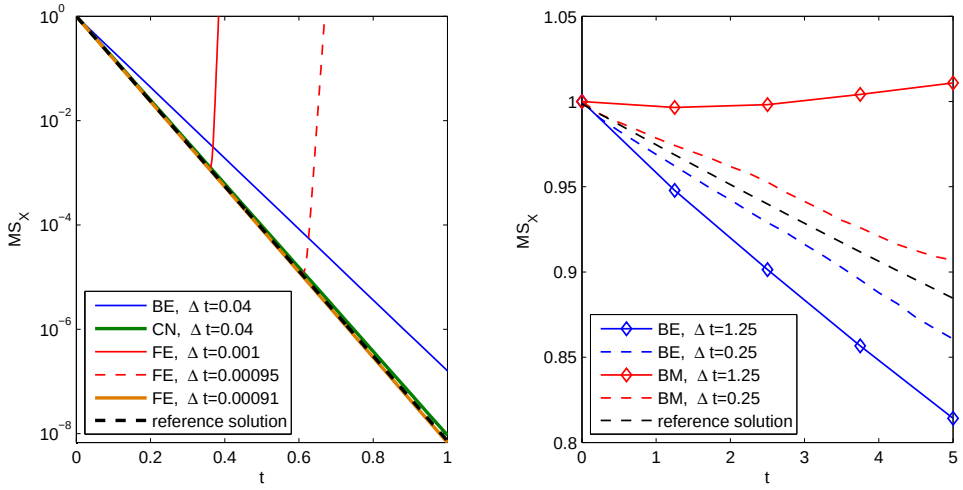
Table 1: Overview of considered methods (rational approximation combined with stochastic integral approximation) applied to the infinite-dimensional geometric Brownian motion. In the second column, the eigenvalues $\Lambda_{k,k}$ defined in (5.3) are given. The last column provides a necessary and sufficient condition in terms of the eigenvalues of $-A$ and Q such that $\rho(\mathcal{S}) < 1$ is fulfilled. Note that Euler–Maruyama is abbreviated by EM.

As it is noted above, the zero solution of (5.1) is asymptotically mean-square stable if and only if the condition $-2\lambda_1 + \mu_1 < 0$ holds. Note that if this condition is fulfilled, then, the numerical approximations based on combinations of an Euler–Maruyama with a backward Euler and a Crank–Nicolson scheme share the mean-square stability properties of the analytical solution without any restriction on V_h or the time step size Δt.

In Figure 4.1(a) we compare the qualitative behaviour of the backward Euler, the Crank–Nicolson, and the forward Euler scheme together with a spectral Galerkin approximation in space and an Euler–Maruyama approximation of the stochastic integral. We choose the dimension of the subspace V_h to be $N_h = 15$. For $i \in \mathbb{N}$, we set the eigenvalues of Q to $\mu_i = i^{-3}$, i.e., $C_\mu = 1$ and $\alpha = 3$. Furthermore, we set $A = \Delta$, i.e., $\nu = 1$. Since $-2\lambda_1 + C_\mu = -2\pi^2 + 1 < 0$, the zero solution of (5.1) is asymptotically mean-square stable.

For the approximation of $\mathbb{E}[\|X_h^j\|_H^2]$ we use a Monte Carlo simulation with $M = 10^6$ independent realizations of $\|X_h^j\|_H^2$, which we compute by using the representation

$$\|X_h(t)\|_H^2 = \sum_{k=1}^{N_h} |x_k(t)|^2. \tag{5.4}$$



(a) Comparison of backward Euler (BE), Crank–Nicolson (CN) and forward Euler (FE) schemes.

(b) Comparison of backward Euler–Maruyama (BE) and backward Euler–Milstein (BM) scheme.

Figure 1: Infinite-dimensional geometric Brownian motion using spectral Galerkin approximations with $N_h = 15$ and different time step sizes Δt.

More specifically, let $\mathrm{MS}_X(t_j) \approx \mathbb{E}[\|X_h^j\|_H^2]$ be given by

$$\mathrm{MS}_X(t_j) = \frac{1}{M} \sum_{i=1}^{M} \sum_{k=1}^{N_h} |\widehat{x}_k^{j,(i)}|^2,$$

where $(\widehat{x}_k^{j,(i)}, i = 1, \dots, M)$ consists of independent samples of numerical approximations of $x_k(t_j)$ with different schemes. Based on Equation (5.4), the reference solution is computed by

$$\mathbb{E}[\|X_h(t)\|_H^2] = \sum_{k=1}^{N_h} \mathbb{E}[|x_k(t)|^2] = \sum_{k=1}^{N_h} \langle X_0, e_k \rangle_H^2 \exp\left((-2\lambda_k + \mu_k)t\right).$$

As it can be seen in Figure 4.1(a), the backward Euler and the Crank–Nicolson scheme reproduce the mean-square stability of the zero solution of of (5.1) already for large time step sizes ($\Delta t = 1/25$), but the forward Euler scheme requires a 44 times smaller Δt in order to reproduce the asymptotic mean-square stability of the zero solution. Here, the finest time step size is given by $\Delta t = 1/1100$ which satisfies the restrictive bound from Table 1 such that $\rho(\mathcal{S}) < 1$. Due to a rapid amplification of oscillations caused by negative values of X_h^j for coarser time step sizes (here, $\Delta t = 1/1000$ and $1/1050$ for which the zero solution is asymptotic mean-square unstable) the mean-square process deviates rapidly from the reference solution at a certain time point.

In Figure 4.1(b) we compare the qualitative behaviour of the backward Euler scheme in combination with an Euler–Maruyama and a Milstein approximation of the stochastic integral on the time interval $t \in [0, 5]$. The parameters $\nu = 8/(5\pi^4)$ (small diffusion) and

$\mu_i = 3/10\, i^{-3}$ are chosen such that the zero solution of the Milstein approximation is asymptotically mean-square unstable for $\Delta t = 1.25$ and asymptotically mean-square stable for $\Delta t = 0.25$. The zero solution of the Euler–Maruyama approximation is for both time step sizes asymptotically mean-square stable. These theoretical results are reproduced in the simulation.

5.2 Galerkin finite element methods

In this section we consider a Galerkin finite element method for the stochastic heat equation with diffusion operator G_2 of Example 4.1. In our context this means that we let V_h be the span of functions that are piecewise linear on an equidistant grid of $[0, 1]$ with N_h interior nodes so that V_h is an N_h-dimensional subspace of $\dot{H}^1$. The refinement parameter is given by $h = (N_h + 1)^{-1}$. With the exception that we now have $U = \dot{H}^1$, all other parameters of the equation are as in Figure 4.1(a) of Section 5.1.

In contrast to the setting in Section 5.1, we can no longer express the solution or its approximation as a sum of one-dimensional geometric Brownian motions and therefore, we no longer have a way to analytically derive necessary and sufficient conditions for $\rho(\mathcal{S}) < 1$. We therefore consider the results of Theorem 4.7 instead. The identity $\lambda_{h,i} = \lambda_i$ for $i \in \mathbb{N}$ does not hold, but an analysis in [19, Section 6.1] shows that $\lambda_{h,i}$ is explicitly given by

$$\lambda_{h,i} = \frac{4\nu}{h^2}\left(\frac{2}{3} + \frac{1}{3}\cos(i\pi h)\right)^{-1} (\sin(i\pi h/2))^2$$

for $i \in \mathbb{N}$. For the convenience of the reader, the sufficient conditions of Theorem 4.7 for the considered approximation schemes are collected in simplified form in Table 2, expressed in terms of stability parameters ρ_{BE}, ρ_{CN} and ρ_{FE}. By introducing the estimate

$$\hat{g} = \left(2\sum_{i=1}^{\infty} \lambda_i^{-1}\right)^{1/2},$$

we replace $\|G_2\|_{L(H;L(U;H))}$ in these conditions by the upper bound derived in Example 4.1. Note that with these parameters, Corollary 4.11 implies the asymptotic mean-square stability of the zero solution of both Equation (5.1) and the backward Euler scheme for the finite element discretization (4.4).

As in Section 5.1 we compare the mean-square behaviour of the backward Euler and the forward Euler schemes in Figure 4.2(a) but now for the finite element discretization up to $T = 2.5$. We observe that the increase of the time step size by a very small amount, i.e., from $\Delta t = 0.00066$ to $\Delta t = 0.00067$, causes the forward Euler system to switch from a stable behaviour to an unstable one. This agrees with our theoretical findings, as ρ_{FE} changes sign in the interval $(0.0066, 0.0067)$, which can be seen in Table 3, so we have theoretically guaranteed stability for the smaller time step but not for the larger one. This indicates that the sufficient condition is sharp in this case.

For the approximation of $\mathbb{E}[\|X_h^j\|_H^2]$, we use the same method as before but take $M =$

rational approximation	$\rho(\mathcal{S}) < 1 \Leftarrow$:
backward Euler	$\rho_{\mathsf{BE}} = \Delta t \operatorname{tr}(Q)\hat{g}^2 - 2\Delta t\lambda_{h,1} - \Delta t^2\lambda_{h,1}^2 < 0$
Crank–Nicolson	$\rho_{\mathsf{CN}} = \max_{k\in\{1,N_h\}} \left\lvert\frac{1-\Delta t\lambda_{h,k}/2}{1+\Delta t\lambda_{h,k}/2}\right\rvert^2 + \frac{\Delta t \operatorname{tr}(Q)\hat{g}^2}{(1+\Delta t\lambda_{h,1}/2)^2} - 1 < 0$
forward Euler	$\rho_{\mathsf{FE}} = \max_{k\in\{1,N_h\}} (1-\Delta t\lambda_{h,k})^2 + \Delta t \operatorname{tr}(Q)\hat{g}^2 - 1 < 0$

Table 2: Overview of temporal approximation schemes applied to the finite element space approximation of the stochastic heat equation with $G = G_2$. The right column shows sufficient conditions for $\rho(\mathcal{S}) < 1$ derived from Theorem 4.7.

10^4 samples in the Monte Carlo approximation. For the computation of the norm in H, we use the fact that with

$$X_h^j = \sum_{m=1}^{N_h} x_m \phi_m,$$

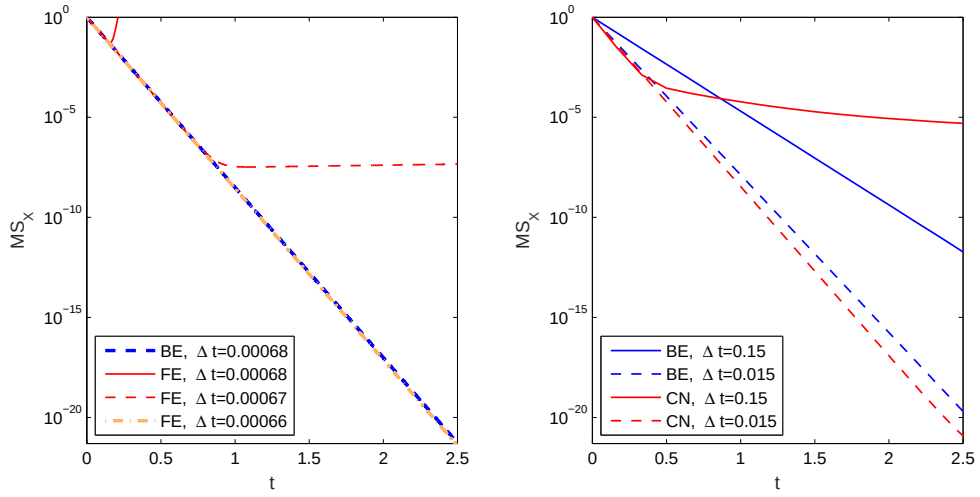
where $x_m \in \mathbb{R}$ for $m = 1, \ldots, N_h$ and $\{\phi_m, m = 1 \ldots, N_h\}$ is the set of *hat functions* that span V_h, we have

$$\|X_h^j\|_H^2 = \sum_{m=1}^{N_h}\sum_{n=1}^{N_h} x_m x_n \langle \phi_m, \phi_n \rangle_H .$$

In Figure 4.2(b) we compare the mean-square behaviour of the the backward Euler scheme and the Crank–Nicolson scheme for $\Delta t = 0.015$ to $\Delta t = 0.15$. Also in this case, we see from Table 3 that ρ_{CN} changes sign when the time step size is increased, which occurs for significantly larger time steps compared to the forward Euler scheme. As one can see from the figure, there is a substantial change in the decay behaviour of $\mathbb{E}[\|X_h^j\|_H^2]$ for the Crank–Nicolson scheme with time step size $\Delta t = 0.15$ compared to $\Delta t = 0.015$, which is no longer convincing to be mean-square stable. Since the sufficient condition $\rho_{CN} < 0$ from Table 2 is not fulfilled for $\Delta t = 0.15$, it is also not clear from the theory that the zero solution of the Crank–Nicolson scheme is asymptotically mean-square stable for such coarse time step sizes.

Properties of Lévy increments

In this appendix we derive properties of the U-valued, square-integrable Lévy process that are used in the proofs of Propositions 4.2 and 4.5. We apply the same setting and notation as in Section 4. For this, we recall that the Karhunen–Loève expansion of L for $t \geq 0$ is



(a) Comparison of backward Euler (BE) and forward Euler (FE) scheme.

(b) Comparison of backward Euler (BE) and Crank–Nicolson (CN) scheme.

Figure 2: Finite Element approximation of the stochastic heat equation with $G = G_2$, $N_h = 15$, and different time step sizes Δt.

Δt	ρ_{BE}	ρ_{CN}	ρ_{FE}
0.15	-5.11613e+00	2.08460e-03	1.99602e+05
0.015	-3.13089e-01	-1.58504e-01	1.91542e+03
0.00068	-1.32387e-02	-1.31050e-02	6.09395e-02
0.00067	-1.30434e-02	-1.29135e-02	3.39709e-04
0.00066	-1.28480e-02	-1.27221e-02	-1.27626e-02

Table 3: Specific values of the stability parameters of Table 2 for varying time step sizes Δt.

given by

$$L(t) = \sum_{k=1}^{\infty} \sqrt{\mu_k} L_k(t) f_k,$$

where $(L_k, k \in \mathbb{N})$ is a sequence of real-valued, uncorrelated, zero-mean Lévy processes that fulfil for all $k \in \mathbb{N}$, $\mathbb{E}[L_k^2(t)] = t$. Here, $(\mu_k, k \in \mathbb{N})$ is the sequence of eigenvalues of the covariance operator Q of L and $(f_k, k \in \mathbb{N})$ is an orthonormal eigenbasis of Q.

Lemma A.1. *Let L be a U-valued Lévy process and let, for $0 \leq a < b$, $\Delta L = L(b) - L(a)$*

and $\Delta t = b - a$*. Then*

$$\mathbb{E}[\Delta L \otimes \Delta L] = \Delta t \sum_{k=1}^{\infty} \mu_k f_k \otimes f_k.$$

Proof. We first note that $\Delta L \otimes \Delta L$ is well-defined as a member of $L^1(\Omega; U^{(2)})$ since

$$\mathbb{E}[\|\Delta L \otimes \Delta L\|_{U^{(2)}}] = \mathbb{E}[\|\Delta L\|_U^2] = \operatorname{tr}(Q)\Delta t < \infty.$$

The increments $\Delta L_k = L_k(b) - L_k(a)$ of the Karhunen–Loève expansion

$$\Delta L = \sum_{k=1}^{\infty} \sqrt{\mu_k} \Delta L_k f_k$$

fulfil $\mathbb{E}[\Delta L_k \Delta L_\ell] = \delta_{k,\ell} \Delta t$ for $k, \ell \in \mathbb{N}$. Thus, we obtain

$$\mathbb{E}[\Delta L \otimes \Delta L] = \sum_{k,\ell=1}^{\infty} \sqrt{\mu_k \mu_l}\, \mathbb{E}[\Delta L_k \Delta L_l]\, f_k \otimes f_\ell = \Delta t \sum_{k=1}^{\infty} \mu_k f_k \otimes f_k. \qquad \square$$

The following lemma is used in Section 4.1 to find a suitable representation of the linear operator $\mathcal{S}$ for the backward Euler–Milstein scheme.

Lemma A.2. *Let* L *be a* U*-valued, square-integrable Lévy process and set for* $0 \le a < b$ *with* $\Delta t = b - a$*,*

$$\Delta^{(2)} L = \sum_{k,\ell=1}^{\infty} \sqrt{\mu_k \mu_\ell} \left(\int_a^b \int_a^s \mathrm{d}L_k(r)\, \mathrm{d}L_\ell(s) \right) f_k \otimes f_\ell \in L^2(\Omega; U^{(2)}).$$

Then

1. $\mathbb{E}\left[\Delta^{(2)} L \otimes \Delta L\right] = 0$*,*
2. $\mathbb{E}\left[\Delta^{(2)} L \otimes \Delta^{(2)} L\right] = \frac{\Delta t^2}{2} \sum_{k,\ell=1}^{\infty} \mu_k \mu_\ell \big((f_k \otimes f_\ell) \otimes (f_k \otimes f_\ell) \big)$*.*

Proof. Since L is stationary, we may assume without loss of generality that $a = 0$ and $b = t > 0$. We first note that

$$\begin{aligned} &\mathbb{E}\left[\left(\int_0^t \int_0^s \mathrm{d}L_i(r)\, \mathrm{d}L_j(s) \right) \left(\int_0^t \int_0^s \mathrm{d}L_k(r)\, \mathrm{d}L_\ell(s) \right) \right] \\ &\qquad = \mathbb{E}\left[\left(\int_0^t L_i(s-)\, \mathrm{d}L_j(s) \right) \left(\int_0^t L_k(s-)\, \mathrm{d}L_\ell(s) \right) \right]. \end{aligned}$$

To simplify this expression, we use the angle bracket process $(\langle X, Y \rangle_t, t \ge 0)$, which for two real-valued semimartingales X and Y with (locally) integrable quadratic covariation $[X, Y]$ is defined as the unique compensator which makes $([X, Y]_t - \langle X, Y \rangle_t, t \ge 0)$ a local martingale. For this, we have the *polarization identity*,

$$\langle X, Y \rangle_t = \frac{1}{4} \left(\langle X + Y, X + Y \rangle_t - \langle X - Y, X - Y \rangle_t \right),$$

which can be found, along with an introduction to this process, e.g., in [27, Section III.5].

For square-integrable martingales M, it holds (see, e.g., [18, Section 8.9]) that

$$\mathbb{E}[\langle M, M\rangle_t] = \mathbb{E}[M^2(t)]$$

and therefore, by the polarization identity, if N is another square-integrable martingale, then,

$$\mathbb{E}[\langle M, N\rangle_t] = \frac{1}{4}\left(\mathbb{E}[(M(t)+N(t))^2] - \mathbb{E}[(M(t)-N(t))^2]\right) = \mathbb{E}[M(t)N(t)].$$

Applying this to the Lévy integral, which is a martingale, we obtain

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^t L_i(s-)\,\mathrm{d}L_j(s)\right)\left(\int_0^t L_k(s-)\,\mathrm{d}L_\ell(s)\right)\right] \\ &= \mathbb{E}\left[\left\langle \int L_i(s-)\,\mathrm{d}L_j(s), \int L_k(s-)\,\mathrm{d}L_\ell(s)\right\rangle_t\right] \\ &= \mathbb{E}\left[\int_0^t L_i(s-)L_k(s-)\,\mathrm{d}\left\langle L_j, L_\ell\right\rangle_s\right],\end{aligned}$$

where the last equality is a property of the angle bracket process and the stochastic integral, see [18, Section 8.9]. Now, if $j = \ell$, we have, since L_j is a Lévy process and $\mathbb{E}[L_j^2(s)] = s$, that $\langle L_j, L_\ell\rangle_s = \langle L_j, L_j\rangle_s = s$ by [26, Chapter 8]. When $j \neq \ell$ on the other hand, L_jL_ℓ is a square-integrable martingale by [26, Theorem 4.49(ii)]. Integration by parts yields

$$[L_j, L_\ell]_s = L_j(s)L_\ell(s) - \int_0^s L_j(r-)\,\mathrm{d}L_\ell(r) - \int_0^s L_\ell(r-)\,\mathrm{d}L_j(r).$$

Therefore, $[L_j, L_\ell]$ is also a square-integrable martingale (with zero mean), because the right hand side is a square-integrable martingale. Since $(\langle L_j, L_\ell\rangle_s, s \geq 0)$ is the *unique* compensator of $[L_j, L_\ell]$ it must follow that $\langle L_j, L_\ell\rangle_s = 0$ for all $s \geq 0$. Thus,

$$\mathbb{E}\left[\int_0^t L_i(s-)L_k(s-)\,\mathrm{d}\left\langle L_j, L_\ell\right\rangle_s\right]$$

is non-zero only if $j = \ell$, and in that case

$$\mathbb{E}\left[\int_0^t L_i(s-)L_k(s-)\,\mathrm{d}\left\langle L_j, L_j\right\rangle_s\right] = \int_0^t \mathbb{E}\left[L_i(s-)L_k(s-)\right]\,\mathrm{d}s.$$

In conclusion we have obtained

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^t\int_0^s \mathrm{d}L_i(r)\,\mathrm{d}L_j(s)\right)\left(\int_0^t\int_0^s \mathrm{d}L_k(r)\,\mathrm{d}L_\ell(s)\right)\right] \\ &= \begin{cases} t^2/2 & \text{for } j=\ell, i=k, \\ 0 & \text{otherwise,}\end{cases}\end{aligned} \tag{A.5}$$

which yields by the monotone convergence theorem that $\Delta^{(2)}L \in L^2(\Omega; U^{(2)})$ with

$$\begin{aligned}\mathbb{E}\left[\|\Delta^{(2)}L\|_{U^{(2)}}^2\right] &= \sum_{k,\ell=1}^{\infty} \mu_k\mu_\ell\, \mathbb{E}\left[\left(\int_0^t \int_0^s \mathrm{d}L_k(r)\,\mathrm{d}L_\ell(s)\right)^2\right] \\ &= \frac{t^2}{2}\sum_{k,\ell=1}^{\infty}\mu_k\mu_\ell = \frac{t^2}{2}\operatorname{tr}(Q)^2 < \infty.\end{aligned}$$

This entails that $\Delta^{(2)}L \otimes \Delta L \in L^1(\Omega; U^{(2)} \otimes U)$, since

$$\begin{aligned}\left(\mathbb{E}\left[\|\Delta^{(2)}L \otimes \Delta L\|_{U^{(2)}\otimes U}\right]\right)^2 &= \left(\mathbb{E}\left[\|\Delta^{(2)}L\|_{U^{(2)}}\|\Delta L\|_U\right]\right)^2 \\ &\le \mathbb{E}\left[\|\Delta^{(2)}L\|_{U^{(2)}}^2\right]\mathbb{E}\left[\|\Delta L\|_U^2\right] = \frac{t^3}{2}\operatorname{tr}(Q)^3 < \infty.\end{aligned}$$

by the Cauchy–Schwarz inequality. Similarly, it holds that

$$\Delta^{(2)}L \otimes \Delta^{(2)}L \in L^1(\Omega; U^{(2)} \otimes U^{(2)}).$$

Therefore, we obtain

$$\mathbb{E}\left[\Delta^{(2)}L \otimes \Delta L\right] = \sum_{k,\ell,m=1}^{\infty} \sqrt{\mu_k\mu_\ell\mu_m}\ \mathbb{E}\left[\Delta L_m \left(\int_0^t\int_0^s \mathrm{d}L_k(r)\,\mathrm{d}L_\ell(s)\right)\right](f_k \otimes f_\ell) \otimes f_m,$$

and, in the same way as the first observation of this proof,

$$\begin{aligned}\mathbb{E}\left[\Delta L_m \left(\int_0^t\int_0^s \mathrm{d}L_k(r)\,\mathrm{d}L_\ell(s)\right)\right] &= \mathbb{E}\left[\left\langle \int \mathrm{d}L_m(s), \int L_k(s-)\,\mathrm{d}L_\ell(s)\right\rangle_t\right] \\ &= \mathbb{E}\left[\int_0^t L_k(s-)\,\mathrm{d}\left\langle L_m, L_\ell\right\rangle_s\right] = 0\end{aligned}$$

This is justified by the observation that $\langle L_m, L_\ell\rangle_s \neq 0$ only if $m = \ell$ and that in this case the expectation of the integral is still zero since L_k has zero expectation.

For the second claim, we note that by (A.5),

$$\begin{aligned}\mathbb{E}\left[\Delta^{(2)}L \otimes \Delta^{(2)}L\right] &= \sum_{i,j,k,\ell=1}^{\infty} \sqrt{\mu_i\mu_j\mu_k\mu_\ell}\big((f_i \otimes f_j) \otimes (f_k \otimes f_\ell)\big) \\ &\qquad \cdot \mathbb{E}\left[\left(\int_0^t\int_0^s \mathrm{d}L_i(r)\,\mathrm{d}L_j(s)\right)\left(\int_0^t\int_0^s \mathrm{d}L_k(r)\,\mathrm{d}L_\ell(s)\right)\right] \\ &= \frac{t^2}{2}\sum_{k,\ell=1}^{\infty}\mu_k\mu_\ell\big((f_k \otimes f_\ell) \otimes (f_k \otimes f_\ell)\big),\end{aligned}$$

which finishes the proof of the lemma. □

References

[1] A. Abdulle and A. Blumenthal. “Stabilized multilevel Monte Carlo method for stiff stochastic differential equations”. In: *J. Comput. Phys.* 251 (2013), pp. 445–460.

[2] L. Arnold. *Stochastic Differential Equations: Theory and Applications*. Translated from the German. John Wiley & Sons, 1974.

[3] A. Barth A.and Lang and C. Schwab. “Multilevel Monte Carlo method for parabolic stochastic partial differential equations”. In: *BIT* 53.1 (2013), pp. 3–27.

[4] A. Barth and A. Lang. “Milstein approximation for advection-diffusion equations driven by multiplicative noncontinuous martingale noises”. In: *Appl. Math. Optim.* 66.3 (2012), pp. 387–413.

[5] A. Barth and A. Lang. “Multilevel Monte Carlo method with applications to stochastic partial differential equations”. In: *Int. J. Comput. Math.* 89.18 (2012), pp. 2479–2498.

[6] A. Barth and A. Lang. “Simulation of stochastic partial differential equations using finite element methods”. In: *Stochastics* 84.2-3 (2012), pp. 217–231.

[7] R. Bellman. “Stochastic transformations and functional equations”. In: *Proc. Sympos. Appl. Math., Vol. XVI*. Amer. Math. Soc., 1964, pp. 171–177.

[8] D. Boffi. “Finite element approximation of eigenvalue problems”. In: *Acta Numer.* 19 (2010), pp. 1–120.

[9] E. Buckwar and C. Kelly. “Towards a systematic linear stability analysis of numerical methods for systems of stochastic differential equations”. In: *SIAM J. Num. Anal.* 48.1 (2010), pp. 298–321.

[10] E. Buckwar and T. Sickenberger. “A structural analysis of asymptotic mean-square stability for multi-dimensional linear stochastic differential systems”. In: *Appl. Numer. Math.* 62.7 (2012), pp. 842–859.

[11] E. Buckwar and R. Winkler. “Multistep methods for SDEs and their application to problems with small noise”. In: *SIAM J. Numer. Anal.* 44.2 (2006), pp. 779–803.

[12] T. Dunst, E. Hausenblas, and A. Prohl. “Approximate Euler method for parabolic stochastic partial differential equations driven by space-time Lévy noise”. In: *SIAM J. Numer. Anal.* 50.6 (2012), pp. 2873–2896.

[13] D. J. Higham. “A-stability and stochastic mean-square stability”. In: *BIT* 40.2 (2000), pp. 404–409.

[14] D. J. Higham. “Mean-square and asymptotic stability of the stochastic theta method”. In: *SIAM J. Numer. Anal.* 38.3 (2000), pp. 753–769.

[15] A. Jentzen and M. Röckner. “A Milstein Scheme for SPDEs”. In: *Found. Comput. Math.* 15.2 (2015), pp. 313–362.

[16] R. Khasminskii. *Stochastic Stability of Differential Equations*. Vol. 66. Stochastic Modelling and Applied Probability. Springer, 2012.

[17] K. Kirchner, A. Lang, and S. Larsson. “Covariance structure of parabolic stochastic partial differential equations with multiplicative Lévy noise”. In: *J. Diff. Equations* (2017).

[18] F. C. Klebaner. *Introduction to Stochastic Calculus with Applications*. Third Edition. Imperial College Press, 2012.

[19] R. Kruse. *Strong and Weak Approximation of Semilinear Stochastic Evolution Equations*. Vol. 2093. Lecture Notes in Mathematics. Springer, 2014.

[20] A. Lang. “A Lax equivalence theorem for stochastic differential equations”. In: *J. Comput. Appl. Math.* 234.12 (2010), pp. 3387–3396.

[21] A. Lang and A. Petersson. “Monte Carlo versus multilevel Monte Carlo in weak error simulations of SPDE approximations”. In: *Mathematics and Computers in Simulation* (2017).

[22] A. Lang and C. Schwab. “Isotropic Gaussian random fields on the sphere: regularity, fast simulation and stochastic partial differential equations”. In: *Ann. Appl. Probab.* 25.6 (2015), pp. 3047–3094.

[23] K. Liu. *Stability of Infinite Dimensional Stochastic Differential Equations with Applications*. Vol. 135. Monographs and Surveys in Pure and Applied Mathematics. Chapman & Hall/CRC, 2006.

[24] G. J. Lord, C. E. Powell, and T. Shardlow. *An Introduction to Computational Stochastic PDEs*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2014.

[25] X. Mao. *Stochastic Differential Equations and Applications*. Horwood Publishing Limited, 2008.

[26] S. Peszat and J. Zabczyk. *Stochastic Partial Differential Equations with Lévy Noise. An Evolution Equation Approach*. Vol. 113. Encyclopedia of Mathematics and Its Applications. Cambridge University Press, 2007.

[27] P. E. Protter. *Stochastic Integration and Differential Equations*. Second Edition. Springer, 2004.

[28] Y. Saito and T. Mitsui. “Stability analysis of numerical schemes for stochastic differential equations”. In: *SIAM J. Numer. Anal.* 33.6 (1996), pp. 2254–2267.

[29] V. Thomée. *Galerkin Finite Element Methods for Parabolic Problems*. 2nd. Vol. 25. Springer Series in Computational Mathematics. Springer, 2006.

[30] M. Wiktorsson. “Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions”. In: *Ann. Appl. Probab.* 11.2 (2001), pp. 470–487.

5 Combining space-time multigrid techniques with multilevel Monte Carlo methods for SDEs

M. Neumüller and A. Thalhammer

In this work we combine multilevel Monte Carlo methods for time-dependent stochastic differential equations with a space-time multigrid method. The idea is to use the space-time hierarchy from the multilevel Monte Carlo method also for the approximation of the solution process by solving the arising linear systems. This symbiosis leads to a robust and parallel method with respect to space, time and probability. We show the performance of this approach by several numerical experiments which demonstrate the advantages of this approach.

Key words. Monte Carlo estimators, Space-time methods, Parallelisation, Multilevel Monte Carlo methods, Multigrid methods, Stochastic differential equations, Finite element methods
AMS subject classifications. 65C05, 60H35, 65Y05, 60H10, 60H15, 65M55

1 Introduction

Stochastic differential equations (SDEs) have become an invaluable tool for modelling time-dependent problems that are perturbed by random influences. Since the importance of such models increases constantly, there is a high demand on improving the efficiency of numerical algorithms for SDEs, especially, if one is interested in the approximation of $\mathbb{E}[\varphi(X(T))]$, where $X(T)$ denotes the (mild) solution of an SDE evaluated at time T.

In this work we focus on approximating $\mathbb{E}[\varphi(X(T))]$ for the solution process of linear SDEs driven by additive noise. For this we combine space-time multigrid methods for approximating solutions of time-dependent deterministic differential equations, see [7] and the references therein, and multilevel Monte Carlo (MLMC) methods, see e.g. [9, 10, 2, 8]. Both methods as such are well-known to be parallelizable, however, the combination of both methods is a completely new approach that enables the full parallelization of the problem in space, time and probability.

The outline of this article is as follows: In Section 2, we introduce two model problems (the Ornstein-Uhlenbeck process and the stochastic heat equation) together with discretization techniques for these model problems with respect to space and time. Afterwards, we consider the multilevel Monte Carlo (MLMC) method for approximating the expectation in Section 3 and we discuss parallelizable space-time multigrid methods based on the inher-

ited space-time hierarchy of the MLMC estimator in Section 4. We conclude by presenting numerical experiments in Section 5.

2 Model problems

Let $T > 0$ and let $(\Omega, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathcal{F}, \mathbb{P})$ be a complete probability space. At first, we consider a one-dimensional model problem given by the stochastic ordinary differential equation (SODE)

$$\begin{aligned} \mathrm{d}u(t) + \lambda u(t)\mathrm{d}t &= \sigma \mathrm{d}\beta(t) \qquad \text{for } t \in (0,T], \\ u(0) &= u_0, \end{aligned} \tag{1}$$

where $\lambda \in \mathbb{R}_0^+, \sigma, u_0 \in \mathbb{R}$ and $\beta = (\beta(t), t \in [0,T])$ is a standard Brownian motion. The solution of this SODE is a special *Ornstein-Uhlenbeck* process defined by

$$u(t) = u_0 e^{-\lambda t} + \sigma \int_0^t e^{-\lambda(t-s)} \mathrm{d}\beta(s), \qquad t \in [0,T]. \tag{2}$$

As second model problem we consider the *stochastic heat equation* on a bounded and convex domain $D \subset \mathbb{R}^d, d = 1,2,3$, with homogeneous Dirichlet boundary conditions. If $d = 2$, then D is assumed to be polygonal and if $d = 3$, then the domain D is polyhedral.

We rewrite the stochastic partial differential equation (SPDE) as a stochastic evolution equation on the Hilbert space $H = L^2(D)$

$$\begin{aligned} \mathrm{d}U(t) &= \mathcal{A}U(t)\mathrm{d}t + G\mathrm{d}W(t) \qquad \text{for } t \in (0,T], \\ U(0) &= U_0 \in \mathcal{D}(\mathcal{A}), \end{aligned} \tag{3}$$

where $G \in L(U;H)$ and $\mathcal{A}v = \Delta_x v$ for $v \in \mathcal{D}(\mathcal{A})$. It is a well-known result that the operator $\mathcal{A}$ is the generator of an analytic semigroup $(S(t), t \in [0,T])$. Subsequently, we denote by $(e_j, j \in \mathbb{N})$ the set of eigenfunctions of the differential operator $\mathcal{A}$, which forms an orthonormal basis of H.

Let U be a separable Hilbert space. Then we assume that $W = (W(t), t \in [0,T])$ is a U-valued Q-Wiener process with a linear, positive definite, symmetric, trace class covariance operator Q. By results from e.g. [5, 12], W can be represented as

$$W(t) = \sum_{j=1}^{\infty} \sqrt{\mu_j} f_j \beta_j(t), \tag{4}$$

where $(\mu_j, j \in \mathbb{N})$ denotes the set of eigenvalues of Q with corresponding eigenfunctions $(f_j, j \in \mathbb{N})$ satisfying $Qf_j = \mu_j f_j$ and $(\beta_j, j \in \mathbb{N})$ is a sequence of independent standard Brownian motions.

Then, by [5], there exists a unique, square-integrable mild solution to SPDE (3)

$$U(t) = S(t)U_0 + \int_0^t S(t-s)G\mathrm{d}W(s) \qquad \text{for } t \in [0,T]. \tag{5}$$

2.1 Discretization of model problems

In this section, we present fully discrete schemes for approximating the solution processes from Eq. (2) and Eq. (5). For this we fix an equidistant partition Θ_K of the time interval $[0,T]$ given by $\Theta_K = \{0 = t_0 < t_1 < \cdots < t_K = T\}$, where for $0 \leq j \leq K$ we choose $t_j = j\Delta t$ with time step size $\Delta t = T/K$.

Discretization of the Ornstein–Uhlenbeck process

For approximating the solution of the Ornstein-Uhlenbeck process (2), we consider the backward Euler–Maruyama scheme given by the recursion

$$(1+\lambda\Delta t)\mathbf{u}_j = \mathbf{u}_{j-1} + \sigma\Delta\beta^j, \qquad \text{for } 1 \leq j \leq K, \tag{6}$$

where $\mathbf{u}_0 = u_0$ and $\Delta\beta^j = \beta(t_j) - \beta(t_{j-1})$. Rewriting the recursion (6) in a matrix-vector representation yields

$$\begin{pmatrix} (1+\lambda\Delta t) & & & \\ -1 & (1+\lambda\Delta t) & & \\ & \ddots & \ddots & \\ & & -1 & (1+\lambda\Delta t) \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_K \end{pmatrix} = \begin{pmatrix} \sigma\Delta\beta^1 + \mathbf{u}_0 \\ \sigma\Delta\beta^2 \\ \vdots \\ \sigma\Delta\beta^K \end{pmatrix}. \tag{7}$$

In this article, we abbreviate this linear system by

$$\mathcal{L}_\tau \mathbf{u} = \mathbf{f}(\omega),$$

where we use the ω-dependency in $\mathbf{f}(\omega)$ to indicate that the right hand side is a random vector.

Discretization of the stochastic heat equation

For the stochastic heat equation we want to obtain a fully discrete approximation U_h^j of the mild solution $U(t_j), t_j \in \Theta_K$, where U_h^j attains values in a finite-dimensional subspace $V_h \subset \mathcal{D}((-\mathcal{A})^{1/2})$. Besides an appropriate time integration method, we apply a discretization scheme in space. For this we consider a standard Galerkin finite element (FE) discretization based on a regular family $(\mathcal{T}_h, h \in (0,1])$ of triangulations of D with maximal mesh size h. Then V_h denotes the space of globally continuous and on $\mathcal{T}_h$ piecewise linear functions. Furthermore, we denote by N_h the dimension of V_h. By using the nodal basis functions $(\phi_i, 1 \leq i \leq N_h)$, the fully discrete approximation scheme based on Galerkin finite elements

in space and on the backward Euler–Maruyama scheme in time is given by (see e.g. [3])

$$(M_h + \Delta t K_h)\mathbf{U}_j = M_h \mathbf{U}_{j-1} + \Delta \mathbf{W}^j \qquad \text{for } 1 \leq j \leq K, \tag{8}$$

where $\Delta \mathbf{W}^j$ denotes the vector representation of the FE approximation of the Q-Wiener increments $G\Delta W^j(\mathbf{x}) = GW(t_j, \mathbf{x}) - GW(t_{j-1}, \mathbf{x}), \mathbf{x} \in D$, and for $j = 0, \ldots, K$,

$$U_h^j = \sum_{i=1}^{N_h} \mathbf{U}_j[i]\phi_i,$$

where $\mathbf{U}_j[i]$ denotes the ith component of the vector $\mathbf{U}_j \in \mathbb{R}^{N_h}$. Here, we denote by M_h the standard *mass* matrix and K_h the standard *stiffness* matrix defined by

$$M_h[i,j] := \int_D \phi_j(\mathbf{x})\phi_i(\mathbf{x})\mathrm{d}\mathbf{x},$$
$$K_h[i,j] := \int_D \nabla\phi_j(\mathbf{x}) \cdot \nabla\phi_i(\mathbf{x})\mathrm{d}\mathbf{x},$$

for $i, j = 1, \ldots, N_h$. Finally, by rewriting the numerical scheme (8) in terms of a matrix-vector formulation we obtain the large linear system

$$\begin{pmatrix} B_h & & & \\ -M_h & B_h & & \\ & \ddots & \ddots & \\ & & -M_h & B_h \end{pmatrix} \begin{pmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \\ \vdots \\ \mathbf{U}_K \end{pmatrix} = \begin{pmatrix} \Delta \mathbf{W}^1 + M_h \mathbf{U}_0 \\ \Delta \mathbf{W}^2 \\ \vdots \\ \Delta \mathbf{W}^K \end{pmatrix}, \tag{9}$$

where $B_h = M_h + \Delta t K_h$. Subsequently we abbreviate the linear system (9) by

$$\mathcal{L}_{h,\tau}\mathbf{U} = \mathbf{F}(\omega).$$

Simulation of Q-Wiener increments

By truncating the Karhunen–Loéve expansion (4) of the Q-Wiener process one can simulate the Q-Wiener increments $(\Delta W^i, i = 1, \ldots, K)$ by defining for $J \in \mathbb{N}$

$$\Delta^J W^i(\mathbf{x}) = P_J \Delta W^i(\mathbf{x}) = \sum_{j=1}^{J} \sqrt{\mu_j} f_j(\mathbf{x}) \Delta\beta_j^i, \qquad \text{for } \mathbf{x} \in D \tag{10}$$

where $\Delta\beta_j^i = \beta_j(t_i) - \beta_j(t_{i-1})$ and P_J denotes the orthogonal projector onto span$(f_1, \ldots, f_J)$. In order to preserve the convergence properties of the backward Euler scheme (8), the truncation parameter J has to be in general coupled to the dimension of the FE space V_h and to the decay rate of the eigenvalues of Q, see e.g. [3, 12].

In general the evaluation of the sum in Eq. (10) is computationally expensive - especially for high-dimensional spaces V_h. However, specific structural properties of the eigenvalues $(f_j, j \in \mathbb{N})$ can be exploited to speed up the simulation of the right hand side $\mathbf{F}(\omega)$. For the

choice of f_j being the eigenvalues of the Laplace operator endowed with periodic boundary conditions on $D = (0,1)^d, d = 2,3$, only a single fast Fourier transform (FFT) is needed to get two independent realisations of the right hand side $\mathbf{F}(\omega)$ on an equidistant grid, see [11, 12]. This has also been implemented in Section 5 for the numerical studies of the stochastic heat equation in higher dimensions (i.e., $D \subset \mathbb{R}^d, d = 2,3$), where for the FFT computations the C++-library $FFTW3$, [6], has been used in the numerical experiments. Furthermore, we want to emphasize that the evaluation of the FFT can also be performed in parallel.

3 Multilevel Monte Carlo methods

The goal is to approximate $\mathbb{E}[\varphi(u(T))]$ or $\mathbb{E}[\varphi(U(T))]$ for a sufficiently smooth mapping $\varphi : H \to B$, where B is a separable Hilbert space, by using suitable estimators. For $Y \in L^2(\Omega; B)$ a common way to approximate $\mathbb{E}[Y]$ is to use a standard *Monte Carlo* (MC) estimator defined by

$$E_M[Y] := \frac{1}{M} \sum_{i=1}^{M} Y^{(i)},$$

where $(Y^{(i)}, i = 1, \dots, M)$ are independent realizations of Y. Here, $L^2(\Omega; B)$ denotes the space of strongly measurable random variables Y that satisfy

$$\|Y\|^2_{L^2(\Omega;B)} := \mathbb{E}[\|Y\|^2_B] < \infty.$$

Due to the rather slow convergence of the MC estimator of order $M^{-1/2}$ in the $L^2(\Omega; B)$-sense, see e.g. [4], the efficient *multilevel Monte Carlo* (MLMC) estimator has been proposed in [9]. For its definition we consider a sequence $(Y_\ell, \ell \in \mathbb{N}_0)$ of approximations of the random variable $Y \in L^2(\Omega; B)$ based on different refinement levels $\ell \in \mathbb{N}_0$ with increasing accuracy and also with increasing computational cost. The MLMC estimator is then given by

$$E^L[Y_L] := \sum_{\ell=0}^{L} E_{M_\ell}[Y_\ell - Y_{\ell-1}],$$

where $Y_{-1} = 0$. The $L^2(\Omega; B)$-error of the MLMC estimator satisfies (see [10])

$$\|\mathbb{E}[Y] - E^L[Y_L]\|_{L^2(\Omega;B)} \leq \|\mathbb{E}[Y - Y_L]\|_B + \left(\sum_{\ell=0}^{L} M_\ell^{-1} \operatorname{Var}[Y_\ell - Y_{\ell-1}] \right)^{1/2} \tag{11}$$

where $\operatorname{Var}[Y] = \mathbb{E}[\|Y - \mathbb{E}[Y]\|^2_B]$ for $Y \in L^2(\Omega; B)$.

3.1 Parameter selection for model problems

We now discuss how to choose the number of samples $(M_\ell, \ell \in \mathbb{N}_0)$ and the refinement parameters h and Δt in order to guarantee the convergence of the MLMC estimator.

Ornstein-Uhlenbeck process

Let u be given in Eq. (2) and for $\ell \in \mathbb{N}_0$ let $\mathbf{u}_{K_\ell}$ be the numerical approximation of $u(T)$ based on the backward Euler–Maruyama scheme (6) with respect to the partition Θ_{K_ℓ} with time step size Δt_ℓ. Furthermore, let $\varphi \in C_b^2(\mathbb{R}, \mathbb{R})$, i.e., $\varphi : \mathbb{R} \to \mathbb{R}$ is twice continuously differentiable with bounded first and second derivatives. Due to the additive noise structure of SDE (1) we obtain by results from [13] that

$$\begin{aligned} |\mathbb{E}[\varphi(u(T)) - \varphi(\mathbf{u}_{K_L})]| &\leq C\Delta t_L, \\ \mathrm{Var}[\varphi(\mathbf{u}_{K_\ell}) - \varphi(\mathbf{u}_{K_{\ell-1}})]^{1/2} &\leq C\Delta t_\ell. \end{aligned}$$

Thus, by similar arguments as in [10], if we choose for any $\varepsilon, C_M > 0$,

$$\begin{aligned} M_0 &= \lceil C_M \Delta t_L^{-2} \rceil, \\ M_\ell &= \lceil C_M \Delta t_\ell^2 \Delta t_L^{-2} \ell^{1+\varepsilon} \rceil \quad \text{for } \ell = 1, \dots, L, \end{aligned} \tag{12}$$

then

$$\mathbb{E}[\varphi(u(T))] - E^L[\varphi(\mathbf{u}_{K_L})]\|_{L^2(\Omega;\mathbb{R})} = \mathcal{O}(\Delta t_L).$$

Stochastic heat equation

Let U be given in Eq. (5) and for $\ell \in \mathbb{N}_0$ let $U_{h_\ell}^{K_\ell}$ be an approximation of $U(T)$ based on the FE backward Euler–Maruyama scheme (8) with respect to the partition Θ_{K_ℓ} and the FE space V_{h_ℓ}. Furthermore, let $\varphi \in \mathcal{C}_b^2(H, B)$, i.e., $\varphi : H \to B$ is twice Fréchet differentiable with bounded first and second Fréchet derivatives. Then by using the results from [1], we get by choosing $\Delta t_\ell = h_\ell^2$ for any $\gamma \in [0, 1)$

$$\begin{aligned} \|\mathbb{E}[\varphi(U(T)) - \varphi(U_{h_L}^{K_L})]\|_B &\leq C h_L^{2\gamma}, \\ \mathrm{Var}[\varphi(U_{h_\ell}^{K_\ell}) - \varphi(U_{h_{\ell-1}}^{K_{\ell-1}})] &\leq C h_\ell^{2\gamma}. \end{aligned}$$

Thus, by [10], if we choose $\Delta t_\ell = h_\ell^2$ and for any $\varepsilon, C_M > 0$,

$$\begin{aligned} M_0 &= \lceil C_M h_L^{-2\gamma} \rceil, \\ M_\ell &= \lceil C_M h_\ell^{2\gamma} h_L^{-2\gamma} \ell^{1+\varepsilon} \rceil \quad \text{for } \ell = 1, \dots, L, \end{aligned} \tag{13}$$

then

$$\|\mathbb{E}[\varphi(U(T))] - E^L[\varphi(U_{h_L}^{K_L})]\|_{L^2(\Omega;B)} = \mathcal{O}(h_L^\gamma).$$

4 Space-time multigrid methods

The idea is to use the space-time hierarchy from the MLMC methods discussed in Section 3.1 also for a space-time multigrid approach. In detail we use the space-time multigrid method presented in [7] to solve the linear system (7) and (9) at once. The advantage is that we can also add parallelization in time direction and for solving the linear system (9) also with respect to the space dimension. So using the space-time hierarchy coming from the MLMC method for the linear solver we obtain an algorithm which can be applied in parallel with respect to space, time and probability. For the space-time multigrid method we use an (inexact) damped block Jacobi smoother, see also [7], i.e. for the problem (7) we use

$$\mathbf{u}^{(n+1)} = \mathbf{u}^{(n)} + \alpha \mathcal{D}_\tau^{-1} \left[\mathbf{f}(\omega) - \mathcal{L}_\tau \mathbf{u}^{(n)} \right] \qquad \text{for } n = 0, 1, \dots,$$

with the diagonal matrix $\mathcal{D}_\tau := \text{diag}(1 + \lambda \Delta t)$. Whereas, for the problem (9) we use the smoothing iteration

$$\mathbf{U}^{(n+1)} = \mathbf{U}^{(n)} + \alpha \mathcal{D}_{h,\tau}^{-1} \left[\mathbf{F}(\omega) - \mathcal{L}_{h,\tau} \mathbf{U}^{(n)} \right] \qquad \text{for } n = 0, 1, \dots,$$

with the block diagonal matrix $\mathcal{D}_{h,\tau} := \text{diag}(B_h)$. To speed up the application of the smoothing procedure we replace the exact inverse of $\mathcal{D}_{h,\tau}$ by applying one iteration of a multigrid V-cycle with respect to the matrix B_h. Moreover we always set the damping parameter to $\alpha = \frac{1}{2}$, see [7] for more details. Choosing $\Delta t \approx h^2$ leads – in combination with the space-time hierarchy coming from the MLMC method – to a robust solver which is independent of the number of time steps K and the time step size Δt.

5 Numerical experiments

In this section we present numerical experiments for the Ornstein–Uhlenbeck process (Section 5.1) and for the stochastic heat equation on $D = (0,1)^d, d = 1, 2, 3,$ (Section 5.2). All results were computed on the RADON1 cluster (a distributed memory cluster located in Linz with 1088 CPU Cores, 8.7TB Memory; consisting of 64 compute nodes each with two 8-core Intel Haswell processors "Xeon E5-2630v3", 2.4Ghz and 128 GB of memory).

5.1 Ornstein-Uhlenbeck process

We consider the SODE (1) with $\lambda = 1, \sigma = 1, T = 1$ and $u_0 = 1$. We want to approximate $\mathbb{E}[\varphi_j(u(T))], j = 1, 2$, where $\varphi_1(x) = x$ and $\varphi_2(x) = x^2$. For both test cases the solution can be computed analytically, where

$$\mathbb{E}[\varphi_1(u(T))] = \mathbb{E}[u(T)] = e^{-T}$$

and

$$\mathbb{E}[\varphi_2(u(T))] = \mathbb{E}[u^2(T)] = \frac{1+e^{-2T}}{2}.$$

For the numerical approximation we consider the backward Euler–Maruyama scheme from Eq. (6) in the matrix-vector representation $\mathcal{L}_\tau \mathbf{u} = \mathbf{f}(\omega)$, which is solved by the time multigrid method described in Section 4. For the approximation of the expectation we consider a multilevel Monte Carlo estimator based on the sample size selection from Eq. (12) with $\varepsilon = \frac{1}{2}$ and $C_M = 10$.

In Table 1, we consider the root mean-square errors of the MLMC method with respect to $\varphi_j, j = 1, 2$, where $\|\mathbb{E}[\varphi_1(u(T))] - E^L[\varphi_1(\mathbf{u}_{K_L})]\|_{L^2(\Omega;\mathbb{R})}$ is approximated by a standard Monte Carlo estimator given by

$$\text{RMS-err}_1 = \left(\frac{1}{M} \sum_{i=1}^{M} \left| \mathbb{E}[u(T)] - E^L[\mathbf{u}_{K_L}]^{(i)} \right|^2 \right)^{1/2},$$

and $\|\mathbb{E}[\varphi_2(u(T))] - E^L[\varphi_2(\mathbf{u}_{K_L})]\|_{L^2(\Omega;\mathbb{R})}$ is approximated by

$$\text{RMS-err}_2 = \left(\frac{1}{M} \sum_{i=1}^{M} \left| \mathbb{E}[u^2(T)] - E^L[\mathbf{u}_{K_L}^2]^{(i)} \right|^2 \right)^{1/2}.$$

Here $(E^L[\varphi_j(\mathbf{u}_{K_L})]^{(i)}, 1 \le i \le M)$ are independent realizations of the MLMC estimator $E^L[\varphi_j(\mathbf{u}_{K_L})]$. For the Monte Carlo simulations we choose $M = 100$ in the numerical experiments from Table 1 and we observe the right convergence behaviour as predicted by the theory.

Table 1: Numerical test for SODE (1) (Ornstein-Uhlenbeck process) - convergence.

L	time steps	M_0	M_L	RMS-err$_1$	EOC	RMS-err$_2$	EOC
0	1	10	10	2.03312E-1	-	1.93142E-1	-
1	2	40	20	1.16614E-1	0.80	1.10059E-1	0.81
2	4	160	50	6.15999E-2	0.92	5.87172E-2	0.91
3	8	640	80	2.99257E-2	1.04	3.06077E-2	0.94
4	16	2560	110	1.52328E-2	0.97	1.70489E-2	0.84
5	32	10240	140	7.75317E-3	0.97	8.50822E-3	1.00
6	64	40960	180	3.71647E-3	1.06	4.33203E-3	0.97
7	128	163840	220	1.81249E-3	1.04	2.27805E-3	0.93
8	256	655360	270	9.65028E-4	0.91	1.26140E-3	0.85
9	512	2621440	310	4.48929E-4	1.10	6.26638E-4	1.01
10	1024	10485760	360	1.96168E-4	1.19	3.42444E-4	0.87

5.2 Stochastic heat equation

For the stochastic heat equation (3) we consider the domain $D = (0,1)^d, d = 1,2,3$, and the initial value

$$U_0(\mathbf{x}) = \prod_{i=1}^{d} \sin(\pi x_i) \qquad \text{for } \mathbf{x} \in D.$$

By choosing $T = 0.1$ and $\varphi(v) = v$ for all $v \in L^2(D)$, we are interested in approximating

$$\mathbb{E}[U(T,\mathbf{x})] = \exp(-d\pi^2 T) \prod_{i=1}^{d} \sin(\pi x_i), \qquad \text{for } \mathbf{x} \in D.$$

For the numerical approximation in space and time, we consider the FE Euler–Maruyama scheme from Eq. (8) on an equidistant spatial triangulation in the matrix-vector formulation $\mathcal{L}_{h,\tau}\mathbf{U} = \mathbf{F}(\omega)$, which is again solved by the space-time multigrid method described in Section 4. For the approximation of the expectation we consider the MLMC method based on the sample size selection (13) with $\varepsilon = \frac{1}{2}$ and $C_M = 10$.

In numerical experiments the root mean-square error $\|\mathbb{E}[U(T)] - E^L[U_{h_L}^{K_L}]\|_{L^2(\Omega;B)}$ is approximated by a standard Monte Carlo estimator, i.e., we consider

$$\text{RMS-err} = \left(\frac{1}{M} \sum_{i=1}^{M} \left\| \mathbb{E}[U(T)] - E^L[U_{h_L}^{K_L}]^{(i)} \right\|_{L^2(D)}^2 \right)^{1/2},$$

where $(E^L[U_{h_L}^{K_L}]^{(i)}, 1 \le i \le M)$ are independent realizations of the estimator $E^L[U_{h_L}^{K_L}]$ and

$$\|\mathbb{E}[U(T)] - E^L[U_{h_L}^{K_L}]^{(i)}\|_{L^2(D)}^2 = \int_D \left| \mathbb{E}[U(T,\mathbf{x})] - E^L[U_{h_L}^{K_L}(\mathbf{x})]^{(i)} \right|^2 \mathrm{d}\mathbf{x}.$$

Stochastic heat equation in 1D

In the one-dimensional case we choose $H = U$ and the eigenvalues of the Q-Wiener process are given by $\mu_j = j^{-(2r+1+\varepsilon)}$ for $r = 2$ and any $\varepsilon > 0$, see e.g. [12] for details. For approximating paths of the Q-Wiener process we use the truncated Karhunen-Loéve expansion from Eq. (4) after the first $J = N_h$ summands, see e.g. [3].

In Table 2 we use $M = 100$ independent realizations of the MLMC estimator and we observe the optimal convergence rates as predicted by the theory. Moreover we give in Table 3 the solving times for one MLMC run for different levels and different distributions of 512 cores, where we share the used cores for the parallelization of the multilevel Monte Carlo simulation and for the parallelization in time for simulating numerical trajectories of (5) by solving the linear system (9). In Table 3 the shortest computation time among each refinement level L with respect to the distribution of the 512 cores is marked in bold. Here we observe that the best possible setting is given by a balanced distribution of cores between parallelization in time and parallelization of the Monte Carlo estimators. For example for

level $L = 7$ the best possible setting is given by 8 cores for time parallelization and 64 cores for the Monte Carlo parallelization.

Table 2: Numerical test for SPDE (3) (stochastic heat equation in 1D) – convergence.

L	time steps	DOF (space)	DOF (ST)	M_0	M_L	RMS-err	EOC
0	1	2	2	10	10	7.83487E-02	-
1	4	4	16	40	20	3.39860E-02	1.20
2	16	8	128	160	30	1.29145E-02	1.40
3	64	16	1024	640	60	5.99035E-03	1.11
4	256	32	8192	2560	90	2.71909E-03	1.14
5	1024	64	65536	10240	120	1.39772E-03	0.96
6	4096	128	524288	40960	150	6.89668E-04	1.02
7	16384	256	4194304	163840	190	3.41996E-04	1.01

Table 3: Numerical test for SPDE (3) (stochastic heat equation in 1D) – computation time with respect to different distributions of 512 cores (in sec).

	cores time / cores Monte Carlo							
L	1 / 512	2 / 256	4 / 128	8 / 64	16 / 32	32 / 16	64 / 8	128 / 4
3	0.04	0.02	0.02	**0.02**	0.03	0.06	0.10	0.14
4	0.27	0.17	**0.12**	0.13	0.16	0.26	0.47	0.93
5	2.64	1.51	**0.95**	1.01	1.17	1.64	2.47	4.41
6	24.12	13.92	13.64	11.47	**10.76**	12.53	15.88	23.50
7	282.46	157.97	153.41	**125.56**	127.84	133.60	146.81	178.76

Stochastic heat equation in 2D and 3D

In this section we consider the stochastic heat equation (3) in higher dimensions, i.e., let $D = (0,1)^d, d = 2,3$. In order to apply the FFT-based simulation techniques of the Q-Wiener increments described in Section 2.1 we consider here the specific choice of eigenvalues

$$\mu_{\mathbf{k}} = \exp(-\mathrm{i}a|\mathbf{k}|^2),$$

where $\mathbf{k} = (k_1, \ldots, k_d)^T \in \mathbb{Z}^d$, $a = 2$ and i denotes the imaginary unit, see [12] for details. The corresponding eigenvalues $f_{\mathbf{k}}$ are given by

$$f_{\mathbf{k}}(\mathbf{x}) = \exp(2\pi\mathrm{i}\langle\mathbf{k}, \mathbf{x}\rangle) \qquad \text{for } \mathbf{x} \in D.$$

For approximating realizations of the Q-Wiener increments we consider a modified version of Eq. (10): For $\mathcal{J} = (J_1, \dots, J_d)^T \in \mathbb{N}^d$ define

$$\Delta^{\mathcal{J}} W^i(\mathbf{x}) = \sum_{\substack{\mathbf{k}\in\mathbb{Z}^d,\\ -J_1/2<|k_1|\leq J_1/2,\\ \dots,\\ -J_d/2<|k_d|\leq J_d/2}} \sqrt{\mu_{\mathbf{k}}} f_{\mathbf{k}}(\mathbf{x}) \Delta\widetilde{\beta}^i_{\mathbf{k}}, \qquad \text{for } \mathbf{x} \in D.$$

Here the increments $\Delta\widetilde{\beta}^i_{\mathbf{k}} = \Delta\beta^{i,\text{real}}_{\mathbf{k}} + \mathrm{i}\Delta\beta^{i,\text{imag}}_{\mathbf{k}}$ are complex, where the real and imaginary parts consist of independent, real-valued Brownian increments $\Delta\beta^{i,\text{real}}_{\mathbf{k}}, \Delta\beta^{i,\text{imag}}_{\mathbf{k}}$, see [12]. In the numerical experiments below we choose $J_i = N_h$ for $i = 1, \dots, d$.

In Tables 4 and 6 we use $M = 100$ realisations to approximate the $L^2(\Omega; H)$-error by a Monte Carlo simulation for the 2D and 3D stochastic heat equation. The experimental outcome illustrates for both test cases that the predicted convergence order $\gamma \approx 1$ is attained.

As for the one-dimensional stochastic heat equation we measure in Tables 5 and 7 the computation times for one run of the MLMC estimator, where we consider different distributions of 512 cores with respect to solving the linear system $\mathcal{L}_{h,\tau}\mathbf{U} = \mathbf{F}(\omega)$ in parallel or for parallelizing the involved Monte Carlo estimators. Similar to the one-dimensional case we obtain that a balanced distribution of cores for the parallelization in time and for the parallelization of the Monte Carlo estimators yields the best computation time. For instance for the numerical experiments in the two-dimensional case at level $L = 6$ the best setting is given by 128 cores for the time parallelization and only 4 cores for the Monte Carlo simulations (844.61 sec). Overall, this results in an approximately 19-times faster solving time than for a straightforward parallelization of the MLMC estimator (16278.1 sec). Another disadvantage of only parallelizing the Monte Carlo estimators is the high amount of memory needed to solve the high dimensional linear systems $\mathcal{L}_{h,\tau}\mathbf{U} = \mathbf{F}(\omega)$ by using the proposed space-time multigrid method. This can be seen in the presented numerical experiments for the case $D = (0, 1)^3$, where we run out of memory (*oom*) on level $L = 5$.

Table 4: Numerical test for SPDE (3) (stochastic heat equation in 2D) – convergence.

L	time steps	DOF (space)	DOF (ST)	M_0	M_L	RMS-err	EOC
1	4	1	4	40	20	3.72225E-02	-
2	16	9	144	160	30	1.85333E-02	1.01
3	64	49	3136	640	60	9.49312E-03	0.97
4	256	225	57600	2560	90	4.65968E-03	1.03
5	1024	961	984064	10240	120	2.10172E-03	1.15
6	4096	3969	16257024	40960	150	1.03803E-03	1.02

Table 5: Numerical test for SPDE (3) (stochastic heat equation in 2D) – computation time with respect to different distributions of 512 cores (in sec).

	cores time / cores Monte Carlo				
L	1 / 512	2 / 256	4 / 128	8 / 64	16 / 32
1	0.07	0.08	0.05	0.07	0.07
2	0.17	0.05	0.08	0.13	0.06
3	0.40	0.22	0.14	**0.12**	0.13
4	6.92	3.59	1.94	1.84	**1.56**
5	115.53	69.83	35.30	34.94	35.31
6	16278.10	7993.41	6310.75	3081.63	1575.11
	cores time / cores Monte Carlo				
L	32 / 16	64 / 8	128 / 4	256 / 2	512 / 1
1	**<0.00**	0.03	0.03	0.01	0.06
2	0.05	**0.03**	0.09	0.13	0.17
3	0.14	0.22	0.42	0.70	1.50
4	1.69	2.07	2.81	5.01	10.11
5	**34.44**	34.68	39.93	50.37	103.75
6	1154.21	953.59	**844.61**	978.95	1168.93

Table 6: Numerical test for SPDE (3) (stochastic heat equation in 3D) – convergence.

L	time steps	DOF (space)	DOF (ST)	M_0	M_L	RMS-err	EOC
1	4	1	4	40	20	1.47372E-02	-
2	16	27	432	160	30	1.23371E-02	0.26
3	64	343	21952	640	60	7.48574E-03	0.72
4	256	3375	864000	2560	90	3.62278E-03	1.05
5	1024	29791	30505984	10240	120	1.76545E-03	1.04

Table 7: Numerical test for SPDE (3) (stochastic heat equation in 3D) – computation time with respect to different distributions of 512 cores (in sec).

	cores time / cores Monte Carlo				
L	1 / 512	2 / 256	4 / 128	8 / 64	16 / 32
1	0.25	0.17	**0.02**	0.05	0.03
2	0.22	0.21	0.14	0.21	0.06
3	2.66	1.36	0.75	**0.36**	0.42
4	133.55	63.82	33.58	32.89	**25.12**
5	*oom*	2886.78	1571.55	1597.29	1501.86
	cores time / cores Monte Carlo				
L	32 / 16	64 / 8	128 / 4	256 / 2	512 / 1
1	0.03	0.03	0.03	0.05	0.05
2	**0.04**	0.07	0.11	0.42	0.37
3	0.56	0.91	1.36	3.57	5.84
4	25.28	27.92	33.79	51.16	99.39
5	1525.51	**1491.09**	1582.12	1671.24	1935.34

6 Conclusions

In this work we combined a space-time multigrid technique for simulating trajectories of the solution process of the Ornstein–Uhlenbeck process and the stochastic heat equation on $D = (0,1)^d, d = 1,2,3$, with a multilevel Monte Carlo estimator to approximate $\mathbb{E}[\varphi(u(T))]$ and $\mathbb{E}[\varphi(U(T))]$, respectively. This combination leads to a fully parallelizable method with respect to space, time and probability.

Special emphasis was laid on constructing the MLMC estimator in such a way that the root mean-square errors converge with order 1 (with respect to the time step size Δt_L for the Ornstein–Uhlenbeck process and with respect to the spatial refinement parameter h_L for the stochastic heat equation). For this the number of numerical realisations $(M_\ell, \ell = 0, \dots, L)$ for the MLMC estimator could be precomputed. In all considered numerical experiments the convergence results could be numerically observed.

For the stochastic heat equation, the robustness of the proposed space-time multigrid solver is guaranteed by [7], since we were using the space-time hierarchy of the MLMC estimator also for the backward Euler–Maruyama approximation of the numerical trajectories, where the space-time coupling satisfies $\Delta t_\ell / h_\ell^2 \approx 1$ on all levels $\ell = 0, \dots, L$.

Furthermore, we investigated the optimal distribution of 512 cores with respect to solving the linear system of the space-time formulation and the parallelization of the involved Monte Carlo estimators. In the presented numerical experiments on the stochastic heat equation it turned out that a balanced distribution of cores between parallelization in time and parallelization of the MLMC estimator provides the best computation times.

References

[1] A. Andersson, R. Kruse, and S. Larsson. “Duality in refined Sobolev-Malliavin spaces and weak approximation of SPDE”. In: *Stoch. Partial Differ. Equ. Anal. Comput.* 4.1 (2016), pp. 113–149.

[2] A. Barth and A. Lang. “Multilevel Monte Carlo method with applications to stochastic partial differential equations”. In: *Int. J. Comput. Math.* 89.18 (2012), pp. 2479–2498.

[3] A. Barth and A. Lang. “Simulation of stochastic partial differential equations using finite element methods”. In: *Stochastics* 84.2-3 (2012), pp. 217–231.

[4] A. Barth, A. Lang, and C. Schwab. “Multilevel Monte Carlo method for parabolic stochastic partial differential equations”. In: *BIT* 53.1 (2013), pp. 3–27.

[5] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and its Applications, 44. Cambridge University Press, Cambridge, 1992.

[6] M. Frigo and S. G. Johnson. “The Design and Implementation of FFTW3”. In: *Proceedings of the IEEE* 93.2 (2005). Special issue on “Program Generation, Optimization, and Platform Adaptation”, pp. 216–231.

[7] M. J. Gander and M. Neumüller. “Analysis of a New Space-Time Parallel Multigrid Algorithm for Parabolic Problems”. In: *SIAM J. Sci. Comput.* 38.4 (2016), A2173–A2208.

[8] M. B. Giles. “Multilevel Monte Carlo methods”. In: *Acta Numer.* 24 (2015), pp. 259–328.

[9] M. B. Giles. “Multilevel Monte Carlo path simulation”. In: *Oper. Res.* 56.3 (2008), pp. 607–617.

[10] A. Lang. “A note on the importance of weak convergence rates for SPDE approximations in multilevel Monte Carlo schemes”. In: *Monte Carlo and quasi-Monte Carlo methods*. Vol. 163. Springer Proc. Math. Stat. Springer, [Cham], 2016, pp. 489–505.

[11] A. Lang and J. Potthoff. “Fast simulation of Gaussian random fields”. In: *Monte Carlo Methods Appl.* 17.3 (2011), pp. 195–214.

[12] G. J. Lord, C. E. Powell, and T. Shardlow. *An Introduction to Computational Stochastic PDEs*. Cambridge Texts in Applied Mathematics. Cambridge University Press, New York, 2014.

[13] G. N. Milstein and M. V. Tretyakov. *Stochastic Numerics for Mathematical Physics*. Scientific Computation. Springer-Verlag, Berlin, 2004.

List of Tables

Introduction

An importance sampling technique in Monte Carlo methods for SDEs with a.s. stable and mean-square unstable equilibrium

Importance sampling techniques for stochastic partial differential equations

Mean-square stability analysis of approximations of stochastic differential equations in infinite dimensions

Combining space-time multigrid techniques with multilevel Monte Carlo methods for SDEs

List of Figures

Curriculum vitae

Andreas Thalhammer

Im Hocheck 12/50
A-4813 Altmünster
Austria
Email: athalhammer@gmx.net

Born 11.06.1991, Gmunden

Education

PhD studies "Technical Sciences" at Johannes Kepler University, Linz
(2015 –2017)
PhD thesis "*Numerical methods for stochastic partial differential equations: Analysis of stability and efficiency*"; advisor Prof. Evelyn Buckwar.

Master studies "Industrial Mathematics" at Johannes Kepler University, Linz
(2013– 2014)
Master thesis: *"Numerical methods for parabolic stochastic partial differential equations"*; advisor Prof. Evelyn Buckwar

Bachelor studies "Technical Mathematics" at Johannes Kepler University, Linz
(2010 – 2013)
Bachelor thesis: "*Konstruktion der Brownschen Bewegung mit Hilfe von Simple Random Walks und deren Anwendung zur Lösung der Wärmeleitungsgleichung*"; advisor Prof. Evelyn Buckwar

Gymnasium "BRG Schloss Traunsee", Altmünster *(2001 – 2009)*

Employment

JKU Linz – Doctoral Program "Computational Mathematics" *(2015-2017)*
Altenberger Straße 69, A-4040 Linz, Research assistant

JKU Linz – Institute for Stochastics *(2014 – 2015)*
Altenberger Straße 69, A-4040 Linz, Teaching assistant

JKU Linz – Institute for Computational Mathematics *(2012 – 2014)*
Altenberger Straße 69, A-4040 Linz, Teaching assistant

Publications

M. Ableidinger, E. Buckwar and A. Thalhammer
"An importance sampling technique in Monte Carlo methods for SDEs with a.s. stable and mean-square unstable equilibrium"
J. Comput. Appl. Math., 316, 3-14 (2017)

E. Buckwar and A. Thalhammer
"Importance sampling techniques for stochastic partial differential equations"
preprint (2016)

A. Lang, A. Petersson and A. Thalhammer
"Mean-square stability analysis of approximations of stochastic differential equations in infinite dimensions"
preprint (2017)

M. Neumüller and A. Thalhammer
"Combining space-time multigrid techniques with multilevel Monte Carlo methods for SDEs"
preprint (2017)